2015

Comparative Population Genomics of Neotropical Forest Birds

Michael Gaston Harvey
Louisiana State University and Agricultural and Mechanical College, mgh272@gmail.com

COMPARATIVE POPULATION GENOMICS OF NEOTROPICAL FOREST BIRDS

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Biological Sciences

by
Michael Gaston Harvey
B.A., Cornell University, 2008
December 2015

“I won't call it rescue, what brought me back here to the old world to drink and decline, and to pretend that the search for another new world was well worth the burn-ing of mine. But sometimes at night, in my dreams, comes the singing of some unknown tropical bird, and I smile in my sleep, thinking Annabelle Lee has finally made it to another new world.”

– Josh Ritter

ACKNOWLEDGEMENTS

I think my advisor, Dr. Robb T. Brumfield, for mentoring and supporting me in many ways, including many I surely am not even aware of. Robb provided me with the flexibility to pursue the scientific questions I was most passionate about and the tools and support to address them. Robb's mentoring style allowed be to thrive, and I could not dream of a better academic advisor.

My committee members, Drs. Jeremy Brown, Michael Hellberg, Donald Baltz, and James Van Remsen, Jr., gave generously of their prodigious knowledge and my research benefitted greatly from their input. Dr. Brown's intimate understanding of phylogenetic methods, Dr. Hellberg's experience in molecular evolution and population genetics, and Dr. Remsen's expertise on Neotropical birds and biology were invaluable assets. Dr. Baltz, my outside committee member, displayed an impressive grasp of the methods and concepts important to my dissertation even they were well outside his area of research.

I owe special thanks to Glenn Seeholzer, with whom I've shared many of my most exciting and inspiring experiences in fieldwork and in academia in general. Glenn has been a first sounding board for most of my (often wild) ideas, has always forced me to examine alternative perspectives, and has driven me to be the best researcher I have the capacity to be. Clare Brown was also a discerning reviewer of my scientific efforts, and tolerated and supported me during some of the most trying episodes of my career. Ryan Terrill and James Maley happily provided both intellectual support and facilitated strategic reprieves from academic pursuits. Brian Smith, Jessica Oswald, John

McCormack, and Liz Derryberry were invaluable role models, collaborators, and friends. Caroline Judy, John Mittermeier, Gustavo Bravo, Andrés Cuervo, Sarah Hird, Marco Rêgo, Glaucia Del-Rio, Oscar Johnson, Rafael Marcondes, Andre Moncrieff, and Santiago Claramunt provided me with an academic family that supported me intellectually and otherwise throughout. César Sanchez, Richard Gibbons, Matt Brady, Vivien Chua, Genevieve Mount, Caleb McMahan, Bill Ludt, Noah Reid, Luke Powell, Brian O'Shea, Dan Lane, Eric Rittmeyer, Sinead Borchert, John McVay, Tara Pelletier, Melissa DeBiasse, Catherine Newman, Jordan Satler, Metha Clock, Even McCulloch, Verity Mathis, Jesús Fernandez, Valerie Derouen, Carlos Prada, Natalia Aristazábal, Sandra Galeano, Emma DeLeon, and many others provided support, advice, and companionship. Rachael Herman provided me with the encouragement and support that made completing my dissertation possible.

Drs. Prosanta Chakrabarty, Mark Hafter, Jacob Esselstyn, and Fred Sheldon were my academic uncles and provided advice and inspiration at many junctures. Drs. Bryan Carstens, Kyle Harms, Philip Stouffer, and Jessica Eberhard, were pseudo-committee-members who were eager to answer questions or discuss research ideas. My undergraduate assistants, Marc Zucker, Claudia Carrizales, Rosalind Remsen, Diana Singh, and Samantha Fields were an invaluable asset and I wish them the best as they pursue their passions.

I thank the many field collectors, curators, and museum staff that provided and cared for many of the materials that were instrumental for my dissertation. I also that the collaborators, field assistants, and indigenous communities that made my field work possible. My research would not have been possible without support from the National

Science Foundation, National Geographic Society, American Ornithologist's Union, Nuttall Ornithological Club, Lewis and Clark Association, Society for Systematic Biology, and Society for the Study of Evolution.

Finally, I want to express my indescribable gratitude to my parents Janice and Neil for inspiring and supporting me from the day I picked up my first pair of binocular through the day that I wrote the last sentence of this dissertation, and to my brother Dennis who was a frequent companion and inspiration throughout.

TABLE OF CONTENTS

ABSTRACT

The causes and implications of differences in geographic variation across species are generally poorly understood, but comparative studies have the potential to provide better understanding of what factors predispose species to undergo population divergence and whether population divergence has lasting evolutionary impacts. Here, I examined geographic variation in birds using molecular data from across the genome. I characterized genetic diversity, estimated population history, and tested for impacts of landscape history as well as ecological traits on genetic parameters. I found evidence that diverse historical processes have led to present-day genetic variation in Neotropical bird species, including divergence, population expansion, migration, and gene flow. Genetic diversity and historical processes differed across species, and some of these differences were associated with habitat. Birds of upland forest had greater genetic diversity, higher divergence between populations, and deeper population histories than birds of floodplain forest in the Amazon. This may result from higher dispersal in floodplain species, recent population expansion in or colonization of floodplain habitats, or persistent demographic differences between habitats. I also found that rates of population divergence within species predicted rates of speciation in their ancestral lineages. This result suggests that traits that predict population divergence within species, such as their habitat associations, will impact their diversification over long evolutionary timescales.

CHAPTER 1: INTRODUCTION

Geographic variation is a pervasive pattern within organisms, and the degree of geographic variation varies widely from species to species. Individuals of some species appear quite different on opposite sides of a river or nearby mountain peaks, whereas individuals in other species look and sound identical on opposite ends of the continent. For example, the Song Sparrow (*Melospiza melodia*) exhibits dramatic morphological differences (Miller 1956) partitioned into 34 subspecies (AOU 1957) and deep genetic structure associated with geography (Zink and Dittmann 1993a) across its distribution in North America. Its relative, the Chipping Sparrow (*Spizella passerina*), has only three weakly differentiated subspecies (AOU 1957) and exhibits no geographic structuring of mitochondrial haplotypes (Zink and Dittmann 1993b). This observation inspires some simple but fundamental questions: How can I compare quantitative measures of geographic variation between different species in an objective manner? Where does geographic variation come from and why does it differ between species? What is the significance of different levels of geographic variation in different species over long evolutionary timescales?

For my dissertation, I explored all of these questions, primarily through the use of molecular approaches and comparative analyses. New genomic approaches and DNA sequencing technologies provided a unique opportunity to gather much larger datasets (Wetterstrand 2015) and investigate more processes and parameters (Carstens et al. 2013) than was possible in the past. However, taking advantage of these tools required some experimentation with different strategies for obtaining data. One such strategy, sequence

capture of ultraconserved elements, involves isolating portions of the genome that are similar across all amniotes (Faircloth et al. 2012). After isolating these regions from many individuals or different species, I can then compare the DNA sequence variation present in the adjoining regions to estimate geographic variation and evolutionary history. Another strategy, restriction associated DNA sequencing (RAD-Seq) involves randomly digesting the genome wherever a certain DNA motif occurs, and then obtaining DNA sequence data from the adjacent regions (e.g. Davey et al. 2011, Elshire et al. 2011). These sequences can then be compared much as those from sequence capture of ultraconserved elements. Pending the widespread availability of whole genomes, these two methods represent promising approaches for obtaining genome-wide DNA sequence data from many individuals and species, and I used both during the course of my dissertation.

All the chapters of my dissertation investigate geographic variation in New World bird species, primarily species of lowland Neotropical forests. Lowland Neotropical birds exhibit variable, but often quite high, levels of geographic genetic variation (Bates 2000, Smith et al. 2015). Moreover, they differ widely in many traits that might impact the development of geographic variation (Parker et al. 1996), and also come from different taxonomic and phylogenetic groups (Remsen et al. 2015) that can serve as semi-independent evolutionary replicates for comparative study.

The four research chapters in my dissertation explored the measurement, sources, and significance of geographic variation based on genomic datasets from New World bird species. In the first chapter, I examined the costs and benefits of two alternative methods for obtaining genomic datasets, sequence capture of ultraconserved elements and RAD-

Seq, using samples from a widespread, lowland Neotropical bird (*Xenops minutus*). In the second chapter, I applied one method (RAD-Seq) to a much larger population-level sample of *Xenops minutus* in order to determine which historical processes have been important in producing geographic variation across its distribution. In the third chapter I also investigated the sources of geographic variation, but I used a comparative framework involving twenty pairs of closely related species that inhabit different habitats, with the goal of determining whether their habitat associations and ecologies determined their levels of geographic variation. This chapter used comparative phylogeographic approaches with data from sequence capture of ultraconserved elements as well as exons. Finally, I explored the significance of geographic variation using a large comparative mitochondrial phylogeographic dataset. In this study, I compared the rate of population differentiation within species to speciation rates inferred from an existing avian phylogeny to determine if variation within species predicted the evolutionary trajectories of their lineages over long timescales.

REFERENCES

American Ornithologists' Union (AOU). 1957. Check-list of North American birds, 5th ed. American Ornithologists' Union, Washington, D.C.

Bates, J. M. 2000. Allozymic genetic structure and natural habitat fragmentation: data for five species of Amazonian forest birds. Condor 102: 770-783.

Carstens, B. C., R. S. Brennan, V. Chua, C. V. Duffie, M. G. Harvey, R. A. Koch, C. D. McMahan, B. J. Nelson, C. E. Newman, J. D. Satler, G. Seeholzer, K. Posbic, D. C. Tank, and J. Sullivan. 2013. Model selection as a tool for phylogeographic inference: an example from the willow Salix melanopsis. Molecular Ecology 22: 4014-4028.

Davey, J. W., P. A. Hohenlohe, P. D. Etter, J. Q. Boone, J. M. Catchen, and M. L. Blaxter. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nature Reviews Genetics 12: 499-510.

Elshire, R. J., J. C. Glaubitz, Q. Sun, J. A. Poland, K. Kawamoto, E. S. Buckler, and S. E. Mitchell. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6: e19379.

Faircloth, B. C., J. E. McCormack, N. G. Crawford, M. G. Harvey, R. T. Brumfield, and T. C. Glenn. 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. Systematic Biology 61: 717-726.

Miller, A. H. 1956. Ecologic factors that accelerate formation of races and species of terrestrial vertebrates. Evolution 3: 262-277.

Parker III, T. A., D. F. Stotz, and J. W. Fitzpatrick. 1996. Ecological and distributional databases. Pages 113-436 in Neotropical Birds: Ecology and Conservation (D. F. Stotz, J. W. Fitzpatrick, T. A. Parker III, and D. Moskovits, eds.). University of Chicago Press, Chicago, Illinois.

Remsen, J. V., Jr., J. I. Areta, C. D. Cadena, A. Jaramillo, M. Nores, J. F. Pacheco, J. Pérez-Emán, M. B. Robbins, F. G. Stiles, D. F. Stotz, and K. J. Zimmer. Version 5 February 2015. A classification of the bird species of South America. American Ornithologists' Union. Available from http://www.museum.lsu.edu/~Remsen/SACCBaseline.html (last accessed September 23, 2015).

Smith, B. T., J. E. McCormack, A. M. Cuervo, M. J. Hickerson, A. Aleixo, C. D. Cadena, J. Pérez-Emán, C. W. Burney, X. Xie, M. G. Harvey, B. C. Faircloth, T. C. Glenn, E. P. Derryberry, J. Prejean, S. Fields, and R. T. Brumfield. 2014. The drivers of tropical speciation. Nature 515: 406-409.

Zink, R. M., and D. L. Dittmann. 1993a. Gene flow, refugia, and evolution of geographic variation in the song sparrow (*Melospiza melodia*). Evolution 47: 717-729.

Zink, R. M., and D. L. Dittmann. 1993b. Population structure and gene flow in the chipping sparrow and a hypothesis for evolution in the genus *Spizella*. The Wilson Bulletin 105: 399-413.

Wetterstrand, M. S. 2015. DNA sequencing costs: Data from the NHGRI genome sequencing program (GSP). Available from http://www.genome.gov/sequencingcosts (last accessed September 23, 2015).

CHAPTER 2: SEQUENCE CAPTURE VERSUS RESTRICTION SITE-ASSOCIATED DNA SEQUENCING FOR PHYLOGEOGRAPHY

INTRODUCTION

New sequencing technologies promise to provide increasingly detailed estimates of species and population histories by resolving rapid radiations (Wagner et al. 2013), improving demographic parameter estimates (Jakobsson et al. 2008), and identifying regions of the genome under selection (Wang et al. 2009). Researchers have recently adopted widely divergent strategies in the approaches used to generate data for systematics. Restriction site associated DNA sequencing (RAD-Seq) is the most widespread method for obtaining genomic datasets from non-model organisms, particularly for studies within species (reviewed in Narum et al. 2013), and these data are increasingly used also for phylogenetics (e.g., Eaton and Ree 2013, Wagner et al. 2013). Sequence capture approaches, typically targeting conserved portions of the genome, have been used primarily for phylogenetics within the field of systematics (e.g., McCormack et al. 2013, Faircloth et al. 2013, Leaché et al. 2014), but these data are also useful for population-level studies (Smith et al. 2014). Other current methods are less applicable to systematics, either because they require high-quality samples for RNA extraction (transcriptomics; Morin et al. 2008), which are poorly represented in genetic resource collections, or because they remain prohibitively expensive when applied to many samples and species (whole genome sequencing; Ellegren 2014, but see Lamichhaney et al. 2015, Nater et al. 2015). Although RAD-Seq and sequence capture are promising tools

for phylogeography and shallow phylogenetic studies of non-model organisms, determining which method is appropriate for particular applications is essential to maximize the benefits of next-generation sequencing to shallow systematics research.

Differences in the potential utility of RAD-Seq and sequence capture stem from a set of issues that impact the resulting datasets. These issues are related to the function and distribution of the loci targeted, the cost of library preparation and sequencing, the assessment of sequence read orthology and locus assembly, the accuracy of variant-calling and genotyping, and the information content within and across resulting loci. Each issue impacts datasets in ways that may bias downstream systematics analyses such as phylogeny reconstruction and demographic parameter estimation (Huang and Knowles 2014, Harvey et al. 2015, Mastretta-Yanes et al. 2014) and all issues may impact the reproducibility of datasets and the comparability of inferences across studies and species. Differences in the effects of each issue between RAD-Seq and sequence capture methods may determine which is preferable for particular applications in phylogeography.

Here, I review the major issues impacting the utility of next-generation sequencing datasets applied to phylogeography studies in non-model species, discuss differences in the importance of each issue relative to RAD-Seq and sequence capture datasets, and discuss how each issue might bias different analyses or applications. Although I focus on the applicability of RAD-Seq and sequence capture to phylogeography, some of the analyses examined are also often employed in population genetics or phylogenetics, and I sometimes use the umbrella term “shallow systematics”. I review existing studies and also re-analyze previously published RAD-Seq and sequence capture datasets from the same population-level samples of a Neotropical bird

(*Xenops minutus*) to provide an empirical example of the differences I observe between methodological approaches. Based on my review and analysis, I also discuss "best practices" for the analysis of these data, and I discuss the appropriateness of both methods for different types of evolutionary studies.

OVERVIEW OF RAD-SEQ AND SEQUENCE CAPTURE

Previous studies have described, in detail, the various strategies for conducting RAD-Seq (e.g., Davey et al. 2011, Elshire et al. 2011, Peterson et al. 2012, Wang et al. 2012, Stolle and Moritz 2013) and sequence capture (e.g., Mamanova et al. 2009, Gnirke et al. 2009, Faircloth et al. 2012, Lemmon et al. 2012, Bi et al. 2012, Hedtke et al. 2013; Li et al. 2013), but a brief review is warranted here. I use RAD-Seq to refer to the family of methods using restriction enzyme digests for genome reduction and high-throughput sequencing, which encompasses many of the methods termed "genotyping by sequencing". RAD-Seq involves digesting genomic DNA with one or more enzymes, adding platform-specific adapters to the fragments, and selecting fragments for sequencing that fall within a particular size distribution (Fig. 1a). This reduces the genome by sampling only those regions near cut sites or where cut sites occur within a certain distance of one another (Baird et al. 2008). Variations on this general method differ primarily in the number of enzymes used (one or two), the types of enzymes used and the frequency of their targeted cut sites, whether random shearing is used on one end, and the approaches used for size selection (Davey et al. 2011, Stolle and Moritz 2013). In most RAD-Seq methods, all fragments from a given locus have at least one static end (the

cut site), meaning that sequence reads are not randomly distributed around a given cut site, which restricts the assembly of longer sequences from RAD-Seq reads (Fig. 1b). Although variations involving paired-end sequencing can produce longer alignments (Willing et al. 2011), most RAD-Seq studies focus on collecting short sequences or single nucleotide polymorphism (SNP) data from groups of short sequences.

Sequence capture involves preparing DNA libraries from randomly fragmented DNA templates and hybridizing these libraries to biotinylated synthetic oligonucleotide probes (60-120 mer) having sequence complementary to hundreds or thousands of genomic regions of interest (Fig. 1c). In the absence of existing genomic resources for a group of interest, probes from genomic regions that are conserved across divergent taxa (e.g., all amniotes) can be used. Streptavidin-coated paramagnetic beads are used to pull down the biotinylated probes and hybridized (target) DNA library fragments, unwanted (non-target) portions of the DNA library are washed away, and targeted fragments are then released from the beads for sequencing (Gnirke et al. 2009; Fisher et al. 2011). Because probes can be tiled across longer regions and enriched fragments are distributed in different positions across targeted loci, reads from sequence capture can be used for assembly of longer sequences (Fig. 1d).

RE-ANALYSIS OF EXISTING DATA

Although an increasing number of both sequence capture and RAD-Seq studies present results pertinent to the issues I describe below, drawing comparisons between

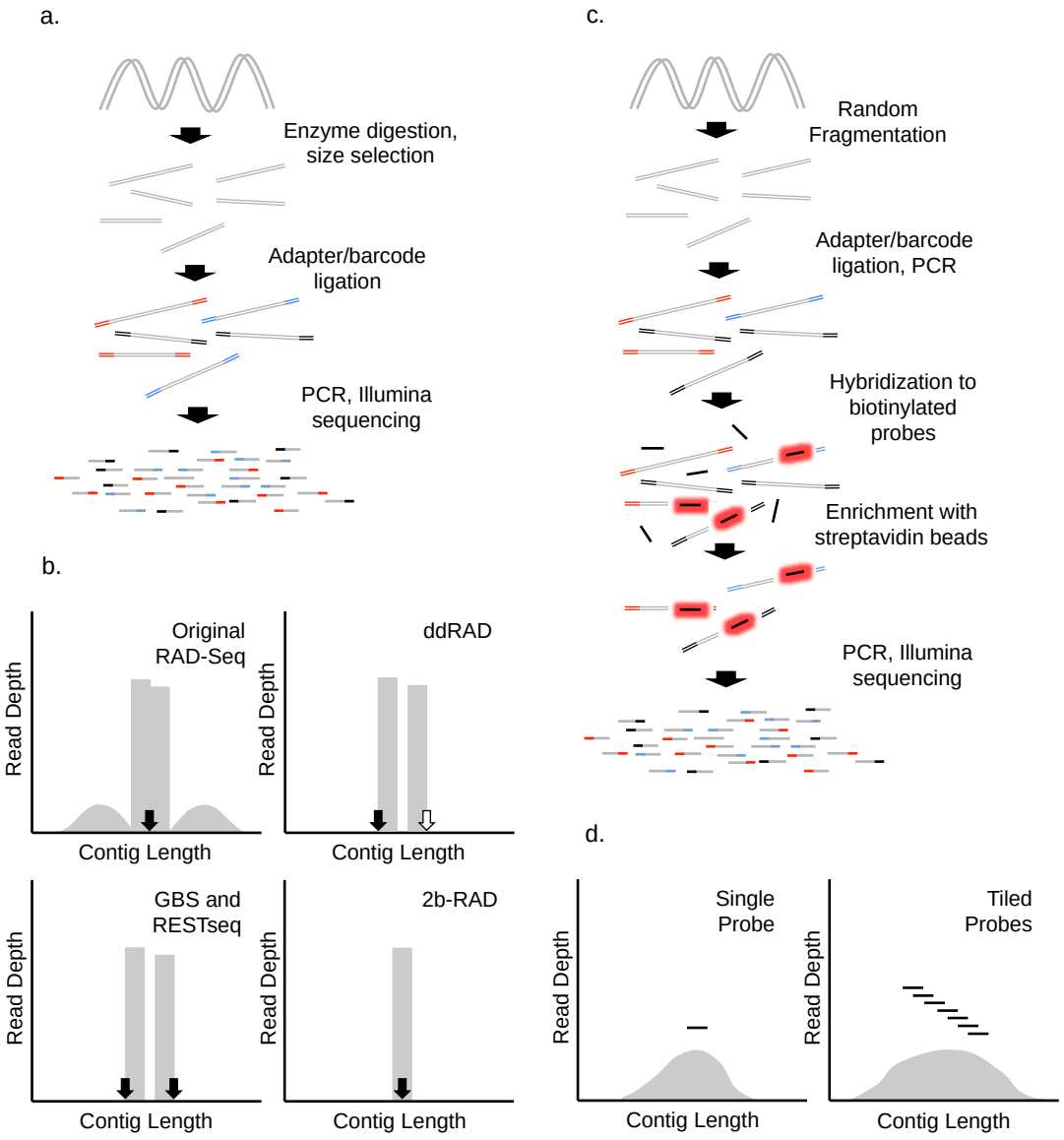


Figure 1. Diagrams of laboratory steps generally required for RAD-Seq (a) and sequence capture (c) as well as typical read distributions from sequencing genomic libraries from various RAD-Seq (b) and sequence capture (d) methods. In (b), enzyme cut sites are depicted using arrows, and different colored arrows represent cut sites for different enzymes.

studies is challenging because they often differ dramatically in sampling and, most importantly, in the methodological decisions made during the process of obtaining and processing sequence data. I therefore supplement my review of existing studies with re-

analysis of some existing RAD-Seq and sequence capture datasets, and I process these data using pipelines that maintain as much consistency as possible between each dataset. Specifically, I analyzed RAD-Seq (Harvey et al. 2015) and sequence capture (Smith et al. 2014) data collected from the same eight individuals of a non-model Neotropical bird (*Xenops minutus*)(Appendix A). Populations of *Xenops minutus* began diverging roughly 5 Mya, and a deep divergence is present between populations on either side of the Andes Mountains (Harvey et al. 2015). I sampled four individuals from populations west of the Andes Mountains and four from populations east of the Andes. I collected RAD-Seq data from all samples using a genotyping by sequencing approach (Elshire et al. 2011), and I collected sequence capture data from ultraconserved elements as described in Faircloth et al. (2012) and Smith et al. (2014). Overall sequencing effort was higher for sequence capture (each sample was one of 44 on an Illumina Hi-Seq lane) than RAD-Seq (each sample was one of 96 on an Illumina Hi-Seq lane) resulting in a mean of 4.96 times higher overall raw read counts in the sequence capture datasets (Appendix A). For RAD-Seq data, I re-processed raw sequence reads and conducted *de novo* assembly using Stacks (Catchen et al. 2011, 2013), and for the sequence capture data I re-processed raw sequence reads using a pipeline described in the seqcap_pop repository (https://github.com/mgharvey/seqcap_pop), which takes advantage of some functions from the Phyluce package (Faircloth 2015). Although the fundamental attributes of RAD-Seq and sequence capture datasets necessitate the use of different methods for dataset assembly, thereby reducing comparability, I used approaches and parameter settings for processing that were as similar as possible between datasets (see Supplemental Information). Although I explored different settings for sequence similarity threshold and

minimum read depth for calling alleles (see below), I conducted most analyses on datasets assembled using a 96% similarity threshold and while requiring 7x minimum read depth per allele (Table 1). I refer to these datasets throughout as the *Xenops minutus* RAD-Seq and sequence capture datasets.

ISSUES IN NEXT-GENERATION SEQUENCING DATASETS

The issues that determine the content of next-generation sequencing datasets are diverse and variable across methods, and I focus here on those issues that I think deserve the greatest weight when selecting RAD-Seq or sequence capture for a project in phylogeography. Differences in how issues impact sequence capture and RAD-Seq datasets are summarized in Table 2.

1) Marker distribution and genomic context

RAD-Seq approaches generally assume there is no genome sequence available for the target organism(s), precluding detailed genomic sampling strategies, and restriction enzymes for RAD-Seq are often selected to cut at sites widely distributed across the genome while avoiding repetitive regions (Elshire et al. 2011). As a result, RAD-Seq sites may come from diverse coding and non-coding regions (Elshire et al. 2011, DaCosta and Sorenson 2014) having potentially heterogeneous genomic contexts and histories, and the sampling of RAD-Seq loci is not truly random, often due to a preponderance of cut sites in regions with particular base compositions (DaCosta and Sorenson 2014).

Table 2. Pros, cons, and applications of RAD-Seq and sequence capture datasets.

Category	RAD-Seq	Sequence Capture
Marker distribution and genomic context	Pro: Dense distribution across genome Con: Anonymous, evolutionary processes largely unknown	Pro: Can be tailored using new genomic information Con: Purifying selection impacts allele frequencies
Practical considerations	Pro: Less expensive, faster	Pro: Works with low-quality samples
Assembly and orthology identification	Pro: Deep coverage, high read overlap	Pro: Over-splitting less problematic
Variant-calling and genotyping	Pro: Fewer rare alleles may make errors easier to distinguish, phasing more straightforward	Pro: Fewer low-coverage rare alleles, no allele dropout
Information content	Pro: More overall information	Pro: More information per locus
Applications	Genome scans, rapid and inexpensive analyses, analyses using species in clades without genomic information, extremely shallow divergences and otherwise intractable relationships.	Comparisons across species, calibrating parameter estimates, targeting loci of known utility or interest, studies using poor-quality samples, studies requiring resolved gene trees, deeper phylogenetic studies.

Sequence capture in non-model species typically targets portions of the genome adjacent to highly conserved regions, such as ultraconserved elements (UCEs; Faircloth et al. 2012) and conserved exons (Bi et al. 2012, Hedtke et al. 2013, Li et al. 2013). Conserved regions are generally selected such that they are distributed widely across available genomes (Faircloth et al. 2012). Ultraconserved elements may serve a structural or regulatory function and may be subject to strong purifying selection (Bejerano et al.

2004; Katzman et al. 2007), while exons are likely to be under several forms of selection, including purifying selection (Ward and Kellis 2012).

In the *Xenops minutus* datasets, Tajima's D is lower (mean = -0.36, SD = 0.82) in sequence capture loci than in RAD-Seq loci (mean = 0.59, SD = 0.90), consistent with the expected effects of purifying selection (Hartl and Clark 2006). I explored the genomic distributions of RAD-Seq and sequence capture loci in *Xenops minutus* by mapping them to the closest genome assembly (*Manacus vitellinus*; Zhang et al. 2014) using Blastn (Altschul et al. 1997). Using stringent alignment settings, 99.4% of UCE loci successfully mapped to the *Manacus* genome compared to 17.7% of the RAD-Seq loci. I used variance in the mean distance between loci across the 92,756 scaffolds in the *Manacus* genome as an index of the level of clustering (Fig. 2). Both ultraconserved elements and RAD-Seq loci recovered from *Xenops minutus* are more clustered than random ($p < 0.001$; Appendix A), but the ultraconserved elements are more clustered than are random subsets of the RAD-Seq loci ($p = 0.001$; Appendix A). The RAD-Seq loci are closer both to predicted protein-coding genes (33.4 +/- 71.3 kbp) and repetitive elements (3.8 +/- 4.9 kbp) than are UCEs (55.0 +/- 84.1 kbp from genes, 4.3 +/- 4.3 kbp from REs). When mapped to a more distant genome (*Taeniopygia guttata*; Warren et al. 2010) with chromosome assemblies available, I found that the proportion of RAD-Seq and sequence capture loci on each chromosome was similar ($R^2 = 0.85$, $p = 2.12 \times 10^{-12}$; Appendix A).

Recent evidence suggests few genomic regions are truly "neutral" (Andolfatto and Przeworski 2000, Schmid et al. 2005), thus examinations of neutral population or species

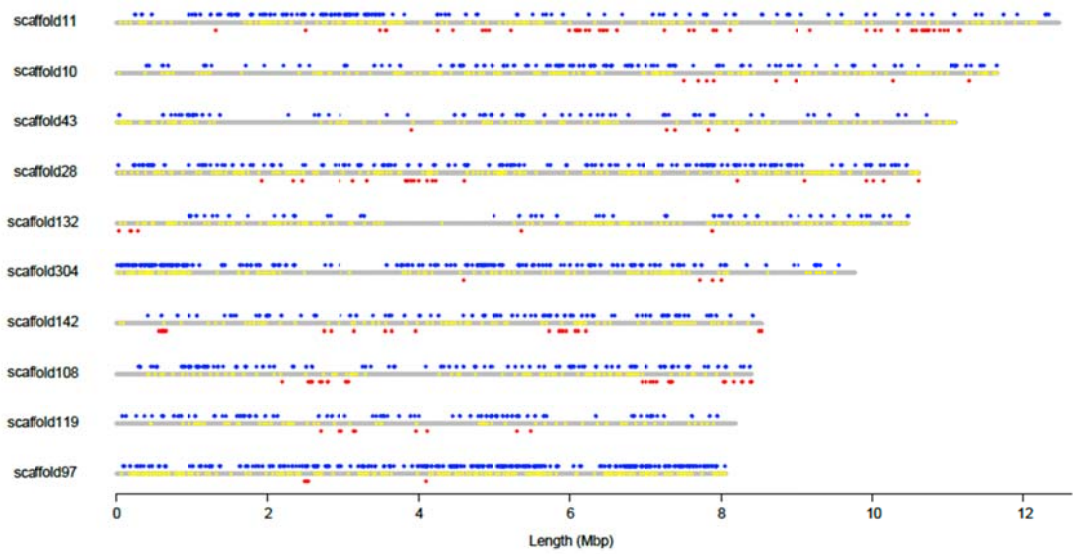


Figure 2. Genomic distributions of RAD-Seq loci (blue dots) and ultraconserved elements (red dots) from *Xenops minutus* when mapped to the ten longest scaffolds in the genome assembly for *Manacus vitellinus*. Yellow spots on the scaffolds are the locations of predicted protein-coding genes.

history may need to account for the action of selection regardless of the loci under examination. Overall, RAD-Seq may better target heterogeneous genomic regions and be more applicable in species in taxonomic groups with little genomic information. Sequence capture is flexible in that probe sets can be augmented or pruned as more genomic information becomes available for a group of interest or as loci are found to be more or less "neutral" or useful for a particular purpose.

2) Practical considerations

Both RAD-Seq and sequence capture can be conducted with relatively small amounts of whole genomic DNA, such as those present in many museum samples. Sequence capture can often be achieved with templates of low concentration or quality (Bi et al. 2012, Guschanski et al. 2013, McCormack et al. in press). Many RAD-Seq

methods require input DNA of higher molecular weight, but some protocols have been developed for samples of poor quality or concentration (e.g., Tin et al. 2014). In addition, sequence capture methods using RAD-Seq libraries as probes may allow RAD loci to be recovered from low-quality samples (Suchan et al. 2015).

Although next-generation sequencing platforms have dramatically reduced the cost and time involved in sequencing (Glenn 2011, Wetterstrand 2015), funding and time may still be limiting in large comparative studies due to expensive library preparations and limitations on the number of samples that can be multiplexed on a single sequencing lane (Harris et al. 2010). The cost of equipment purchase is negligible because both RAD-Seq and sequence capture can be conducted using equipment that is standard in most molecular labs (Gnirke et al. 2009, Elshire et al. 2011), although a sonicator is necessary for some sequence capture protocols. Sequence capture is generally more expensive than RAD-Seq due to the costs associated with more involved library preparation and purchasing enrichment probes. For my *Xenops minutus* datasets, sample preparation and sequencing for RAD-Seq datasets cost roughly $40 US per sample and sequence capture datasets cost roughly $60 per sample. Sequence capture may also require greater sequencing depth (to get sufficient coverage of more variable regions flanking conserved probe targets) and thus higher sequencing cost than RAD-Seq on a per locus basis, but this is offset because sequence capture approaches target informative, single-copy loci more efficiently. In the *Xenops minutus* dataset, 90.1% of raw reads were on-target and included in the assembly, versus only 44.1% of the RAD-Seq reads (Appendix A).

Similarly, time investment is modest for both methods (Gnirke et al. 2009; Elshire et al. 2011), although sequence capture is slower due to the additional hybridization and enrichment steps. For about one hundred samples, library preparation for RAD-Seq can be completed in about two days, whereas an equivalent number of sequence capture libraries can be prepared in two to four days. Commercial library preparation and sequencing services, requiring only quantified whole genomic DNA, are available for both RAD-Seq and sequence capture.

3) Assembly and orthology identification

In next-generation sequencing workflows, the process of dataset assembly is non-trivial, and its success depends on the attributes of the reads coming off the sequencer as well as the methodological decisions made during bioinformatics processing. Assembling reads into sequences and aligning them across individuals into loci is a critical component of processing next-generation sequencing datasets and has received the most attention, particularly in prior studies of the utility of RAD-Seq data for systematics (e.g. Rubin et al. 2012, Cariou et al. 2013). A primary initial concern in orthology assessment of next-generation sequence reads was whether, in divergent lineages separated by millions of years of evolutionary history, reads could be reliably recovered from sufficient loci for historical inference. It is now clear that, even in less conserved regions such as those potentially targeted by RAD-Seq protocols, sufficient orthologous data can be recovered for population-level analyses and phylogenetic analyses involving species with divergences of up to 60 My or more (Rubin et al. 2012, Cariou et al. 2013).

A secondary issue is whether the process of orthology assessment introduces biases in the resulting datasets that impact downstream analyses. Interactions between sequence divergence and the assembly parameters selected during data processing can have profound impacts on resulting datasets. Many assembly programs are available (e.g., Zerbino and Birney 2008, Simpson et al. 2009, Catchen et al. 2011) and all use sequence similarity, in some form, to assemble reads. Reads with high sequence similarity are expected to come from the same locus and are assembled, whereas those with low similarity are expected to come from different loci and are not (Pop and Salzberg 2008, Chaisson et al. 2009). A threshold is used to determine which reads belong to a single locus, but variation in genetic divergence across the genome and among study systems makes determination of an appropriate threshold challenging (Ilut et al. 2014, Harvey et al. 2015b). If the similarity threshold applied is too low, reads from different loci will be assembled into a single locus and treated as orthologous ("under-splitting"), whereas if the threshold is too high, alleles belonging to a single locus may be split into separate alignments ("over-splitting").

The use of similarity thresholds for assembly is a concern for both RAD-Seq and sequence capture datasets. Under-splitting may be frequent in RAD-Seq datasets if enzyme cut sites in different genomic regions fall within similar sequences, however previous results from simulated and empirical RAD-Seq data suggest that under-splitting is infrequent (Ilut et al. 2014) and does not introduce enough signal to impact downstream analyses (Rubin et al. 2012). In many sequence capture approaches, loci are vetted to ensure they are single-copy in existing genome sequences (e.g., Faircloth et al. 2012), but the possibility of paralogous reads assembling to these loci in other taxa exists.

That said, high sequence similarity within conserved regions may permit easier discrimination between orthologous and paralogous reads in sequence capture datasets, and the biology of ultraconserved elements suggests that paralogy is low (Derti et al. 2006).

We examined the relative frequency of under-splitting in RAD-Seq and sequence capture datasets from *Xenops minutus*. Examining raw assemblies, I used the number of alignments containing individuals with three or more alleles (birds are diploid) as an index of the frequency of putative paralogy (Ilut et al. 2014, Harvey et al. 2015b). I found that under-splitting is of roughly equal and low (<0.6% of loci) prevalence in both RAD-Seq and sequence capture datasets assembled under a range of similarity thresholds (Appendix A). The under-split loci were identified and easily removed from both datasets. These results suggest under-splitting and paralogy are a relatively minor concern for both RAD-Seq and sequence capture datasets, at least in species without highly repetitive genomes and when examining relatively recently diverged samples that do not necessitate the use of liberal similarity thresholds.

Over-splitting may be frequent in short read datasets when high similarity among reads is required for assembly (Ilut et al. 2014). In *de novo* RAD-Seq assembly, over-splitting results in the separation of alternative alleles at a locus into separate alignments. Conversely in sequence capture datasets, because reads are being aligned to a sequence determined *a priori*, over-splitting results in the loss of reads and therefore alleles that are highly divergent from the reference. High similarity thresholds for locus assembly, such as 98 or 99%, are often used with short read datasets (e.g., Catchen et al. 2011, Lu et al. 2013), potentially aggravating the issue of over-splitting. The net result of over-splitting

in both RAD-Seq and sequence capture datasets is a decrease in the mean number of alleles within loci. I explored the frequency of over-splitting in RAD-Seq and sequence capture using the datasets from *Xenops minutus*. I used the loss of alleles at a high similarity threshold (99%) relative to a lower similarity threshold (94%) as an index of the prevalence of over-splitting. I found that using a stringent similarity threshold resulted in an average loss of 19.4% of alleles in the RAD-Seq dataset, but only 6.9% of alleles in the sequence capture dataset (Fig. 3a). Over-splitting may be more severe in the RAD-Seq dataset both because of greater divergence among alleles within RAD-Seq loci relative to ultraconserved elements and because each over-split locus results in two less variable alignments in RAD-Seq data. In sequence capture, conversely, over-splitting results in only one less variable locus because reads are aligned to a sequence that is determined *a priori*. Although using less stringent similarity thresholds for assembly can alleviate the impact of over-splitting (Ilut et al. 2014, Harvey et al. 2015b), RAD-Seq datasets may be more sensitive to this key assembly parameter. High conservation and low paralogy in sequence capture loci may improve discrimination of orthologous versus paralogous reads and be more amenable to assembly under low similarity thresholds. Correctly assessing orthology reduces bias in parameter estimates within studies (Mastretta-Yanes et al. 2014), and improves the comparability of datasets and inferences across studies (Harvey et al. 2015b).

4) Variant calling and genotyping

Calling variants and genotyping individuals is the next important step after assembly when processing next-generation sequencing data, and this process is equally fraught with potential issues. Short read sequencer errors introduce spurious nucleotides or indels that may be identified as alleles if they are not correctly vetted (Dohm et al. 2008). Sequencing errors are problematic in both sequence capture and RAD-Seq datasets. The impact of sequencing errors on a dataset can potentially be reduced both by using filters and by calling alleles in a probabilistic framework (Nielsen et al. 2011).

Sequence read depth and evenness of sequence read depth across alleles are perhaps the most critical pieces of information researchers can use to distinguish true alleles from errors. Thus, differences in read depth or evenness across alleles between sequence capture and RAD-Seq may impact the relative success of variant calling between the two methods. Sequence capture and many RAD-Seq approaches require polymerase chain reaction (PCR) to obtain sufficient template for sequencing, and PCR can result in amplification bias and inconsistent coverage across alleles (Aird et al. 2011). Read depth in sequence capture datasets is often higher in the conserved regions that the probe targets than in the more variable flanking regions (Fig. 1c), which are critical for calling variants. RAD-Seq datasets may also exhibit high variability in read depth across loci or amplification bias between alleles that decreases the evenness of coverage (DaCosta and Sorenson 2014). In both methods, PCR cycles should be reduced as much as possible to reduce amplification bias, and PCR duplicate reads should be removed during bioinformatics processing. PCR duplicates are more difficult to remove

bioinformatically in RAD-Seq datasets because duplicates are detected by scanning for overlapping reads, but overlapping reads are expected in most RAD-Seq approaches.

We assessed the frequency of putative errors in RAD-Seq and sequence capture data from *Xenops minutus* by examining the relative read depth across rare (singleton) SNP alleles I identified in the alignments. As expected, I found that a low read depth filter (requiring 3x coverage per allele) resulted in larger datasets (Appendix A), but a low read depth filter resulted in more singleton alleles than assemblies requiring higher coverage (11x)(Appendix A). The RAD-Seq dataset, however, was more impacted by the read depth filter I applied: I recovered 8.0 times as many singletons at 3x depth than I recovered at 11x depth, compared to only 4.6 times as many singletons at 3x versus 11x depth in the sequence capture dataset (Fig. 3b). This suggests that a high proportion of singleton alleles in my RAD-Seq dataset had low coverage and may represent spurious allele calls. It is unclear if variability in read depth across alleles is particular to my RAD-Seq dataset, or if protocols could be optimized to reduce coverage bias, for example, by reducing the number of loci targeted in the RAD-Seq protocol.

Aside from sequencing errors, other artifacts can be observed in the allele frequency spectrum and can potentially be removed at the variant calling stage. Any lingering paralogous data present in an assembly (see above) can potentially be vetted during the variant calling process. High heterozygosity is typically attributed to paralogy because it may reflect the inclusion of sequences from two divergent loci in a single alignment (Hohenlohe et al. 2011). Paralogs can be removed by filtering for heterozygosity (although this can also remove highly variable loci or loci under diversifying selection) or for loci with higher than expected read depth. Allele dropout

due to restriction site polymorphisms is a different problem that may result in elevated homozygosity because individuals that would be heterozygous appear as homozygotes (but see Gautier et al. 2013), and it is unique to RAD-Seq datasets. Within recently diverged species and species with small effective population sizes, allele dropout may not be severe, but it is likely to increase in datasets including multiple species or deeply diverged populations (DaCosta and Sorenson 2014).

The spectrum of expected allele frequencies in a set of markers impacts the ability to detect artifacts. Rare alleles representing errors may be more difficult to identify in conserved loci targeted by sequence capture because I expect a high proportion of rare alleles under purifying selection (Hartl and Clark 2006). Conversely, loci containing paralogous reads resulting in high heterozygosity may be easier to distinguish in conserved loci if there is lower overall heterozygosity in these regions.

Examining allele frequency spectra from the *Xenops minutus* datasets reveals patterns that may be due to the artifacts mentioned above and to real differences between RAD-Seq and sequence capture loci (Fig. 3c). The conserved loci recovered from sequence capture had higher overall frequencies of singleton alleles than the RAD-Seq loci (48% from sequence capture vs. 22% from RAD-Seq using the 7x coverage threshold; Fig. 3c), consistent with the action of purifying selection. In RAD-Seq, 77% of genotypes were homozygous versus 56% of sequence capture genotypes in the *Xenops minutus* dataset, and the proportion of loci deficient in heterozygotes relative to Hardy-Weinberg expectations was slightly higher in the RAD-Seq (61%) than the sequence capture dataset (55%). This discrepancy may be due to a greater impact of allele dropout,

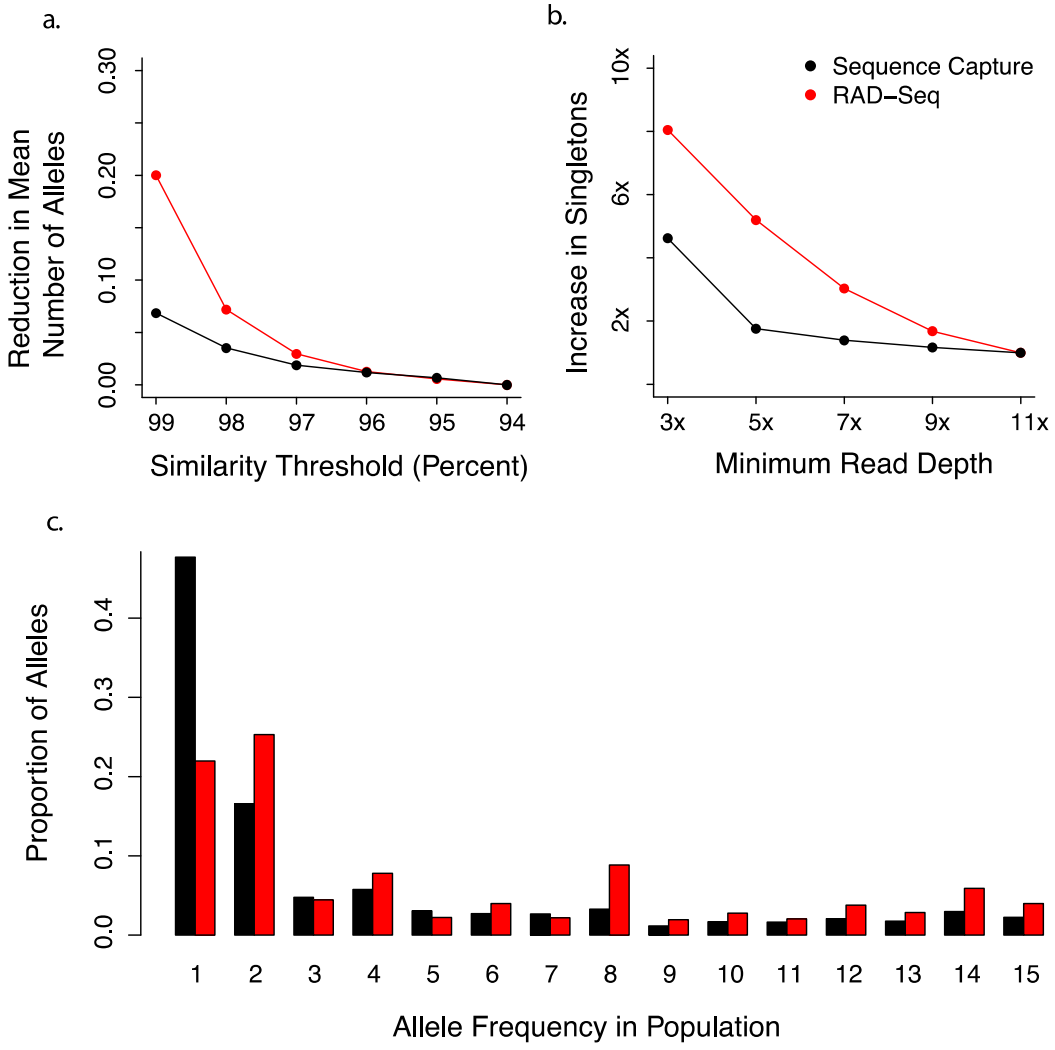


Figure 3. (a) The reduction in alleles in sequence capture and RAD-Seq datasets when using stringent similary thresholds for assembly. (b) The increase in singleton alleles (potential errors) in RAD-Seq and sequence capture datasets at lenient minimum read depth thresholds for alleles. (c) Frequency spectra of all alleles in *Xenops minutus* sequence capture and RAD-Seq datasets processed using a 96% similarity threshold and requiring 7x read depth per allele.

PCR bias, or uneven sequencing coverage in the RAD-Seq dataset, or it may be a result of real genotype frequency differences between the sets of markers.

Phasing alleles is a final important element in variant calling when researchers need to reconstruct haplotypes. In single-end RAD-Seq alignments, alleles are easily phased based on whether they occur on the same reads or not (read-backed phasing). In paired-end RAD-Seq and sequence capture, however, reads are not entirely overlapping and phasing may require additional methods. Read-backed phasing can be used for sites in close proximity or in datasets with paired-end reads, but probabilistic models are required for sites in which read-backed phasing is not possible. These models use information from panels of reference individuals sampled previously or from other individuals in the dataset to impute the most probable combinations of alleles for heterozygous individuals. Model-based phasing introduces an extra step, and potentially additional estimation error, in datasets from paired-end RAD-Seq and sequence capture.

5) Information content

Although RAD-Seq and sequence capture both result in much greater informative variation than equivalent investment in older Sanger sequencing methods, RAD-Seq generally results in greater total aligned sequence and more informative characters. The information, however, is partitioned into shorter loci. In *Xenops minutus*, for example, I assembled 158,329 RAD-Seq loci averaging 95.6 (SD = 0.62) bp in length, whereas for sequence capture I obtained 1,358 loci averaging 590 (SD = 209) bp in length (Table 1). The total number of segregating sites for RAD-Seq (213,740) was much higher than for sequence capture (5,524), but the mean number of segregating sites per locus was higher for sequence capture: 4.07 (SD = 3.57) versus 1.35 (SD = 1.56). RAD-Seq may be

Table 1. Summary of *Xenops minutus* dataset attributes.

	Sequence Capture	RAD-Seq
Number of Loci	1,358	158,329
Mean Locus Length (sd)	590.36 (209.21)	95.55 (0.62)
Mean Number of Segregating Sites (sd)	4.07 (3.57)	1.35 (1.56)
Mean Number of Alleles (sd)	4.52 (2.88)	2.04 (1.14)
Mean Watterson's Θ (sd)	0.0021 (0.0017)	0.0057 (0.0065)
Mean Tajima's D (sd)	-0.36 (0.82)	0.59 (0.90)

preferable for estimating challenging parameters, at least in recently diverged samples, because the greater number of polymorphisms increases the chances of finding a shared allele on a very short phylogenetic branch or representing a rare migration event. For approaches requiring more information per locus, such as analyses based on gene tree estimation, sequence capture may be preferable.

POTENTIAL IMPACTS OF BIASES ON SYSTEMATIC INFERENCES

The issues described above may shape datasets in ways that make them more or less appropriate or biased for downstream systematic analyses. Sequence capture and RAD-Seq datasets yield broadly concordant results for phylogenetic analyses at deep timescales, depending on the steps used for dataset assembly (Leaché et al. 2015), but their relative utility for different systematics analyses applied to study recently diverged populations or species is largely unexplored. Genome-wide scans, for example to identify signatures of selection or gene flow, are often conducted using RAD-Seq loci due to their dense distribution across the genome (Hohenlohe et al. 2010). Conserved regions targeted

by sequence capture may be insufficiently dispersed across the genome for use in genome-wide scans. The short length of many RAD-Seq loci, however, makes mapping them to divergent genomes challenging, thus RAD-Seq may not be appropriate for identifying the genomic context of outlier loci in species without available genome assemblies. As with many markers, RAD-Seq loci may come from heterogeneous genomic regions impacted by diverse neutral and non-neutral processes, so scans will need to account for alternative explanations for outlier loci or migrant alleles.

Demographic inference may be affected by the distribution of allele frequencies in a dataset. Purifying selection on conserved regions may leave signatures, such as an excess of rare alleles, that complicate estimation of neutral demographic histories. Heterozygote deficiencies in RAD-Seq datasets may also impact estimates of demographic parameters including θ and admixture. I estimated demographic parameters using a model-based approach in BP&P (Yang and Rannala 2010) with both RAD-Seq and sequence capture data from *Xenops minutus*. The demographic model included two daughter populations comprising the four samples from west of the Andes Mountains and the four samples east of the Andes Mountains, both of which diverged from a common ancestral population. I compared estimates of effective population size by normalizing the divergence time estimates from RAD-Seq and sequence capture datasets. I found that effective population sizes in the daughter populations were similar between datasets (Appendix A), but the estimate of ancestral effective population size was lower from sequence capture than from RAD-Seq data (Fig. 4b). This discrepancy is likely due to the high frequency of rare alleles restricted to a single population in the sequence capture

alignments that, although perhaps a result of purifying selection, also fits a history of expansion in those populations.

Phylogenetic tree estimation may be complicated if allele loss results in a downward bias in the mutational spectrum (Huang and Knowles 2014). This bias may produce shallower gene trees and lower genetic distances, particularly between the most divergent individuals in a sample (Harvey et al. 2015b). I examined branch lengths from *Xenops minutus* trees inferred using BUCKy (Larget et al. 2010), which are estimated in coalescent units based on quartet concordance factors for each branch. As observed in prior studies (Leaché et al. 2015), internal branch lengths in trees estimated from RAD-Seq data were short relative to those estimated from sequence capture data in *Xenops minutus*, perhaps as a result of the loss of the most divergent alleles (Fig. 4c,d). Heterozygote deficiencies in RAD-Seq datasets may impact genetic distances and branch lengths from some phylogenetic methods. The lengths of the terminal branches observed in BUCKy trees for *Xenops minutus* are determined by the gene trees from loci in which individuals are homozygous for rare alleles. These branch lengths are longer in the RAD-Seq tree than the sequence capture tree, consistent with the high levels of homozygosity ohserved in the RAD-Seq dataset. Despite differences in phylogenetic branch lengths, relative genetic distances among individuals were highly correlated between RAD-Seq and sequence capture *Xenops minutus* datasets (CADM test coefficient of concordance = 0.935, $p < 0.001$, Fig. 4a), suggesting that allele frequency differences between the datasets did not strongly influence distance estimates.

Both RAD-Seq and, to a lesser extent, sequence capture loci have low per-locus information content relative to many traditional markers. Low per-locus information

content complicates analyses that depend on accurate parameter estimates from individual loci. It may be challenging to fit models of molecular evolution to RAD-Seq loci due to their low information content, and poorly resolved gene trees may complicate analyses such as gene tree-species tree estimation (Lanier et al. 2014). Concordance analysis of gene trees from RAD-Seq and sequence capture in *Xenops minutus* using BUCKy (Larget et al. 2010) revealed that consensus relationships were supported by relatively few loci (Fig. 4c,d). Most gene trees contained polytomies as a result of low information content in alignments. Concordance was lower among RAD-Seq loci than among sequence capture loci, presumably due to the lower resolution of RAD-Seq gene trees. The consensus trees inferred for both datasets from all loci were topologically identical, however (Fig. 4c,d). The phylogenetic utility of conserved loci is still under debate (e.g., Betancur R. et al. 2013, Salichos and Rokas 2013). Methods that successfully integrate across the small amounts of information present in many loci, including methods that examine independent SNPs (e.g. Gutenkunst et al. 2009, Bryant et al. 2012), may be desirable for sequence capture and particularly RAD-Seq datasets.

The large datasets produced by RAD-Seq and sequence capture raise computational concerns. Although the sizes of both RAD-Seq and sequence capture datasets can be tailored according to researcher needs, RAD-Seq datasets are generally larger. Depending on the question being addressed, very large datasets may not be needed and additional data may unnecessarily complicate analyses (Davey et al. 2011). Conversely, evolutionary events that are difficult to estimate may require large amounts of data to address, and larger datasets also offer the ability to subsample loci informing a research question *post-hoc*. To take advantage of the information in large datasets,

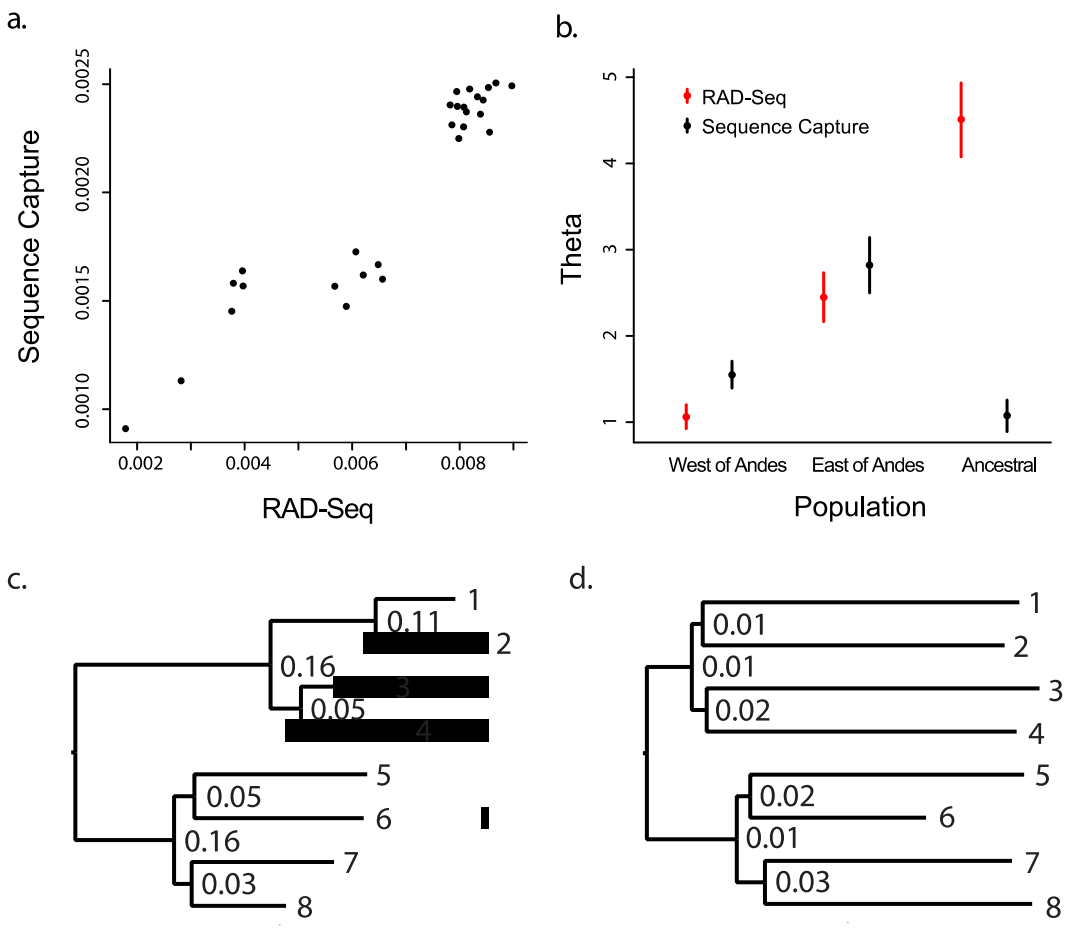


Figure 4. Impacts of dataset biases on inferences from systematic analyses of *Xenops minutus* data from RAD-Seq and sequence capture. (a) Relative pairwise Jukes-Cantor corrected distances between individuals, (b) mutation-scaled effective population size (θ) estimates for daughter and ancestral populations, (c) BUCKy tree from sequence capture and (d) BUCKy tree from RAD-Seq, with node values representing the number of gene trees from that dataset containing each clade.

computationally demanding methods, such as full likelihood phylogeny estimation, may have to take a back seat to faster, summary methods (e.g., Liu et al. 2009, Larget et al. 2010, Chaudhary et al. 2014).

COMPARING ACROSS DATASETS AND CALIBRATING PARAMETERS

The same RAD-Seq loci often cannot be recovered across divergent species due to mutations at restriction sites (Rubin et al. 2012). Studies have successfully recovered some shared loci at moderately deep (~60 My) timescales in *Drosophila* (Rubin et al. 2012, Cariou et al. 2013), but sequence capture is more effective for recovering the same loci at even very deep timescales (up to about ~400 My; Faircloth et al. 2012, Faircloth et al. 2013). When identical loci cannot be recovered in different RAD-Seq studies, comparability across species relies on the assumption that RAD-Seq loci in each species represent a random sample from the genome. RAD-Seq loci, however, provide a biased sample of the genome that is dependent on the restriction enzyme selected and base composition of the genome under study (DaCosta and Sorenson 2014). Furthermore, locus assembly is not random with respect to the level of genetic variation and genome complexity in the species being examined. Over-splitting may disproportionately impact species with higher divergence (Huang and Knowles 2014, Harvey et al. 2015b), whereas under-splitting may be a greater issue in species with repetitive genomes (Ilut et al. 2014, Harvey et al. 2015b). Methods are available for informed selection of assembly parameters in order to reduce the impacts of over-splitting and under-splitting (e.g. Ilut et al. 2014, Harvey et al. 2015b), but whether these will be sufficient to permit comparability across species is unclear. Sequence capture of loci containing conserved regions appears to be the safer option for obtaining genomic data if datasets or inferences are to be directly compared across divergent species.

Within species, estimation of real values for demographic and historical parameters requires calibrating genetic diversity or substitution rates. For species lacking fossil data or divergences tied to dated geological events, standardized mutation rates must be adopted from studies of related species. Standardized mutation rates, however, can only be implemented if the same loci are examined and if datasets are comparable. Calibration, therefore, may be quite challenging in RAD-Seq datasets from non-model species. Calibration across species is possible in sequence capture datasets, however, if datasets are assembled and variants called in the same way and if the alignments are trimmed such that they contain the same sites.

CONCLUSIONS

Although I observe broad concordance in RAD-Seq and sequence capture datasets and resulting inferences, the differences I observed suggest that they are not equally useful for different shallow systematics applications. RAD-Seq is the fastest and least expensive means to obtain large amounts of data, and its application to single-species population studies, genome scans, groups without genomic information, and species with very shallow histories will surely continue to grow. Additional research, however, should focus on understanding the evolutionary processes impacting RAD-Seq loci across the genome, reducing the loss of alleles during assembly and variant-calling, and integrating across the low information content in many short loci. Sequence capture may be preferable for obtaining intraspecific datasets that are comparable across species, calibrating parameter estimates for demographic or phylogenetic studies, tailoring marker

sets to target genomic regions of interest, incorporating low-quality samples, or conducting studies at deeper timescales. The potential biases introduced by purifying selection in conserved genomic regions, however, require continued investigation.

We anticipate that the issues associated with sequence capture and RAD-Seq will change as the methods evolve and improve. Moreover, new methods are sure to appear and existing methods such as whole-genome sequencing and re-sequencing will become more affordable in the near future. Even with these new and improved methods, many of the issues I have described will continue to require attention as genomic approaches and next-generation sequencing datasets become prevalent in phylogeography and systematics as a whole.

REFERENCES

Aird, D., M. G. Ross, W.-S. Chen, M. Danielsson, T. Fennell, C. Russ, D. B. Jaffe, C. Nusbaum, and A. Gnirke. 2011. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. Genome Biology 12: R18.

Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and J. D. Lipman. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Research 25: 3389-3402.

Andolfatto, P. and M. Przeworski. 2000. A genome-wide departure from the standard neutral model in natural populations of *Drosophila*. Genetics 156: 257-268.

Baird, N. A., P. D. Etter, T. S. Atwood, M. C. Currey, A. L. Shiver, Z. A. Lewis, E. U. Selker, W. A. Cresko, and E. A. Johnson. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One 3: e3376.

Bejerano, G., M. Pheasant, I. Makunin, S. Stephen, W. J. Kent, J. S. Mattick, and D. Haussler. 2004. Ultraconserved elements in the human genome. Science 304: 1321-1325.

Betancur-R., R., G. Naylor, and G. Orti. 2013. Conserved genes, sampling error, and phylogenomic inference. Systematic Biology 63: 257-262.

Bi, K., D. Vanderpool, S. Singhal, T. Linderoth, C. Moritz, and J. Good. 2012. Transcriptome-based exon capture enables highly cost-effective comparative genomic data collection at moderate evolutionary scales. BMC Genomics 13: 403.

Bryant, D., R. Bouckaert, J. Felsenstein, N. A. Rosenberg, and A. RoyChoudhury. 2012. Inferring species trees directly from biallelic genetic markers: Bypassing gene trees in a full coalescent analysis. Molecular Biology and Evolution 29: 1917-1932.

Catchen J. M., A. Amores, P. A. Hohenlohe, W. A. Cresko, and J. H. Postlethwait. 2011. Stacks: building and genotyping loci de novo from short-read sequences. G3 Genes Genomes Genetics 1: 171-182.

Catchen J. M., P. A. Hohenlohe, S. Bassham, A. Amores, and W. A. Cresko. 2013. Stacks: an analysis tool set for population genomics. Molecular Ecology 22: 3124-3140.

Cariou, M., L. Duret, and S. Charlat. 2013. Is RAD-seq suitable for phylogenetic inference? An in silico assessment and optimization. Ecology and Evolution 3: 846-852.

Chaisson, M. J., D. Brinza, and P. A. Pevzner. 2009. De novo fragment assembly with short mate-pair reads: Does the read length matter? Genome Research 19: 336-346.

DaCosta, J. M. and M. D. Sorenson. 2014. Amplification biases and consistent recovery of loci in a double-digest RAD-seq protocol. PLoS One 9: e106713.

Chaudhary, R., D. Fernández-Baca, and J. G. Burleigh. 2014. MulRF: A software package for phylogenetic anlaysis using multi-copy gene trees. Bioinformatics 31: 432-433.

Davey, J. W., P. A. Hohenlohe, P. D. Etter, J. Q. Boone, J. M. Catchen, and M. L. Blaxter. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nature Reviews Genetics 12: 499-510.

Derti, A., F. P. Roth, G. M. Church, and C.-T. Wu. 2006. Mammalian ultraconserved elements are strongly depleted among segmental duplications and copy number variants. Nature Genetics 38: 1216-1220.

Dohm, J. C., C. Lottaz, T. Borodina, and H. Himmelbauer. 2008. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. Nucleic Acids Research 36: e105.

Eaton, D. A., and R. H. Ree. 2013. Inferring phylogeny and introgression using RADseq data: An example from flowering plants (Pedicularis: Orobanchaceae). Systematic Biology 62: 689-706.

Ellegren, H. 2014. Genome sequencing and population genomics in non-model organisms. Trends in Ecology and Evolution 29: 51-63.

Elshire, R. J., J. C. Glaubitz, Q. Sun, J. A. Poland, K. Kawamoto, E. S. Buckler, and S. E. Mitchell. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6: e19379.

Faircloth, B. C. 2015. PHYLUCE is a software package for the analysis of conserved genomic loci. Bioinformatics doi:10.1093/bioinformatics/btv646.

Faircloth, B. C., J. E. McCormack, N. G. Crawford, M. G. Harvey, R. T. Brumfield, and T. C. Glenn. 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. Systematic Biology 61: 717-726.

Faircloth, B. C., L. Sorenson, F. Santini, and M. E. Alfaro. 2013. A phylogenomic perspective on the radiation of ray-finned fishes based upon targeted sequencing of ultraconserved elements (UCEs). PLoS One 8: e65923.

Fisher, S., A. Barry, J. Abreu, B. Minie, J. Nolan, T. M. Delorey, G. Young, T. J. Fennell, A. Allen, and L. Ambrogio. 2011. A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. Genome Biology 12: R1.

Gautier, M., K. Gharbi, T. Cezard, J. Foucaud, C. Kerdelhué, P. Pudlo, J.-M. Cornuet, and A. Estoup. 2013. The effect of RAD allele dropout on the estimation of genetic variation within and between populations. Molecular Ecology 22: 3165-3178.

Glenn, T. C. 2011. Field guide to next-generation DNA sequencers. Molecular Ecology Resources 11: 759-769.

Gnirke, A., A. Melnikov, J. Maguire, P. Rogov, E. M. LeProust, W. Brockman, T. Fennell, G. Giannoukos, S. Fisher, C. Russ, S. Gabriel, D. B. Jaffe, E. S. Lander, and C. Nusbaum. 2009. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nature Biotechnology 27: 182-189.

Guschanski, K., J. Krause, S. Sawyer, L. M. Valente, S. Bailey, K. Finstermeier, R. Sabin, E. Gilissen, G. Sonet, Z. T. Nagy, G. Lenglet, F. Mayer, and V. Savolainen. 2013. Next-generation museomics disentangles one of the largest primate radiations. Systematic Biology 62: 539-554.

Gutenkunst, R. N., R. D. Hernandez, S. H. Williamson, and C. D. Bustamante. 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genetics 5: 1-11.

Harris, J. K., J. W. Sahl, T. A. Castoe, B. D. Wagner, D. D. Pollock, and J. R. Spear. 2010. Comparison of normalization methods for construction of large, multiplex amplicon pools for next-generation sequencing. Applied Environmental Microbiology 76: 3863-3868.

Hartl, D. L. and A. G. Clark. 2006. Principles of Population Genetics: Fourth Edition. Sinauer, Sunderland, Massachusetts, Sinauer.

Harvey, M. G. and R. T. Brumfield. 2015. Genomic variation in a widespread Neotropical bird (*Xenops minutus*) reveals divergence, population expansion, and gene flow. Molecular Phylogenetics and Evolution 83: 305-316.

Harvey, M. G., C. D. Judy, G. F. Seeholzer, J. M. Maley, G. R. Graves, and R. T. Brumfield. 2015. Similarity thresholds used in DNA sequence assembly from short reads can reduce the comparability of population histories across species. PeerJ 3: e895.

Hedtke, S. M., M. J. Morgan, D. C. Cannatella, and D. M. Hillis. 2013. Targeted enrichment: maximizing orthologous gene comparisons across deep evolutionary time. PLoS One 8: e67908.

Hohenlohe, P. A., S. Bassham, P. D. Etter, N. Stiffler, E. A. Johnson, and W. A. Cresko. 2010. Population genomics of parallel adaptation in threespine stickleback using sequences RAD tags. PLoS Genetics 6: e1000862.

Hohenlohe, P. A., S. J. Amish, J. M. Catchen, F. W. Allendorf, and G. Luikart. 2011. Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. Molecular Ecology Resources 11: 117-122.

Huang, H. and L. L. Knowles. 2014. Unforeseen consequences of excluding missing data from next-generation sequences: Simulation study of RAD sequences. Systematic Biology doi:10.1093/sysbio/syu046.

Ilut, D. C., M. L. Nydam, and M. P. Hare. 2014. Defining loci in restriction-based reduced representation genomic data from nonmodel species: Sources of bias and diagnostics for optimal clustering. BioMed Research International 2014: 675158.

Jakobsson, M., S. W. Scholz, P. Scheet, J. R. Gibbs, J. M. VanLiere, H.-C. Fung, Z. A. Szpiech, J. H. Degnan, K. Wang, and R. Guerreiro. 2008. Genotype, haplotype and copy-number variation in worldwide human populations. Nature 451: 998-1003.

Katzman, S., A. D. Kern, G. Bejerano, G. Fewell, L. Fulton, R. K. Wilson, S. R. Salama, and D. Haussler. 2007. Human genome ultraconserved elements are ultraselected. Science 317: 915-915.

Lamichhaney, S., J. Berglund, M. S. Allmén, K. Maqbool, M. Grabherr, A. Martinez-Barrio, M. Promerová, C.-J. Rubin, C. Wang, N. Zamani, B. R. Grant, P. R. Grant, M. T. Webster, and L. Andersson. 2015. Evolution of Darwin's finches and their beaks revealed by genome sequencing. Nature 518: 371-375.

Lanier, H. C., H. Huang, and L. L. Knowles. 2014. How low can you go? The effects of mutation rate on the accuracy of species-tree estimation. Molecular Phylogenetics and Evolution 70: 112-119.

Larget, B. R., S. K. Kotha, C. N. Dewey, and C. Ané. 2010. BUCKy: Gene tree/species tree reconciliation with Bayesian concordance analysis. Bioinformatics 26: 2910-2911.

Leaché, A. D., P. Wagner, C. W. Linkem, W. Böhme, T. J. Papenfuss, R. A. Chong, B. R. Lavin, A. M. Bauer, S. Nielsen, E. Greenbaum, M-O. Rödel, A. Schmitz, M. LeBreton, I. Ineich, L. Chirio, E. A. Eniang, S. Baha El Din, A. R. Lemmon, and F. T. Burbrink. 2014. A hybrid phylogenetic-phylogenomic approach for species tree estimation in African Agama lizards with applications to biogeography, character evolution, and diversification. Molecular Phylogenetics and Evolution 79: 215-230.

Leaché, A. D., A. S. Chavez, L. N. Jones, J. A. Grummer, A. D. Gottscho, and C. W. Linkem. 2015. Phylogenomics of Phrynosomatid lizards: Conflicting signals from sequence capture versus restriction site associated DNA sequencing. Genome Biology and Evolution 7: 706-719.

Lemmon, A. R., S. A. Emme, and E. M. Lemmon. 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. Systematic Biology 61:727-744.

Li, C., M. Hofreiter, N. Straube, S. Corrigan, and G. J. Naylor. 2013. Capturing protein-coding genes across highly divergent species. Biotechniques 54: 321-326.

Liu, L., L. Yu, D. K. Pearl, and S. V. Edwards. 2009. Estimating species phylogenies using coalescence times among sequences. Systematic Biology 58: 468-477.

Lu, F., A. E. Lipka, J. Glaubitz, R. Elshire, J. H. Cherney, M. D. Casler, E. S. Buckler, and D. E. Costich. 2013. Switchgrass genomic diversity, ploidy, and evolution: Novel insights from a network-based SNP discovery protocol. PLoS Genetics 9: e1003215.

Mamanova, L., A. J. Coffey, C. E. Scott, I. Kozarewa, E. H. Turner, A. Kumar, E. Howard, J. Shendure, and D. J. Turner. 2010. Target-enrichment strategies for next-generation sequencing. Nature Methods 7: 111-118.

Mastretta-Yanes, A., N. Arrigo, N. Alvarez, T. H. Jorgensen, D. Piñero, and B. C. Emerson. 2014. Restriction site-associated DNA sequencing, genotyping error estimation and *de novo* assembly optimization for population genetic inference. Molecular Ecology Resources 15: 28-41.

McCormack, J. E., W. L. E. Tsai, and B. C. Faircloth. In press. Sequence capture of ultraconserved elements from bird museum specimens. Molecular Ecology Resources doi:10.111/1755-0998.12466.

McCormack, J. E., M. G. Harvey, B. C. Faircloth, N. G. Crawford, T. C. Glenn, and R. T. Brumfield. 2013. A phylogeny of birds based on over 1,500 loci collected by target enrichment and high-throughput sequencing. PLoS One 8: e54848.

Morin, R. D., M. Bainbridge, A. Fejes, M. Hirst, M. Krzywinski, T. J. Pugh, H. McDonald, R. Varhol, S. J. Jones, and M. A. Marra. 2008. Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. Biotechniques 45: 81.

Nater, A., R. Burri, T. Kawakami, L. Smeds, and H. Ellegren. 2015. Resolving evolutionary relationships in closely related species with whole-genome sequencing data. Systematic Biology 64: 1000-1017.

Narum, S. R., C. A. Buerkle, J. W. Davey, M. R. Miller, and P. A. Hohenlohe. 2013. Genotyping-by-sequencing in ecological and conservation genomics. Molecular Ecology 22: 2841-2847.

Nielsen, R., J. S. Paul, A. Albrechtsen, and Y. S. Song. 2011. Genotype and SNP calling from next-generation sequencing data. Nat. Rev. Genet. 12: 443-451.

Peterson, B. K., J. N. Weber, E. H. Kay, H. S. Fisher, and H. E. Hoekstra. 2012. Double digest RADseq: An inexpensive method for *de novo* genotyping in model and non-model species. PLoS One 7: e37135.

Pop, M., and S. L. Salzberg. 2008. Bioinformatics challenges of new sequencing technology. Trends in Genetics 24: 142-149.

Rubin, B. E., R. H. Ree, and C. S. Moreau. 2012. Inferring phylogenies from RAD sequence data. PLoS One 7: e33394.

Salichos, L., and A. Rokas. 2013. Inferring ancient divergences requires genes with strong phylogenetic signals. Nature 497: 327-331.

Schmid, K. J., S. Ramos-Onsins, H. Ringys-Beckstein, B. Weisshaar, and T. Mitchell-Olds. 2005. A multilocus sequence survey in *Arabidopsis thaliana* reveals a genome-wide departure from a neutral model of DNA sequence polymorphism. Genetics 169: 1601-1615.

Simpson, J. T., K. Wong, S. D. Jackman, J. E. Schein, S. J. Jones, and I. Birol. 2009. ABySS: A parallel assembler for short read sequence data. Genome Research 19: 1117-1123.

Smith, B. T., M. G. Harvey, B. C. Faircloth, T. C. Glenn, and R. T. Brumfield. 2014. Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. Systematic Biology 63: 83-95.

Springer, M. S. and J. Gatesy. 2015. The gene tree delusion. Molecular Phylogenetics and Evolution 94: 1-33.

Stolle, E., and R. F. A. Moritz. 2013. RESTseq: Efficient benchtop population genomics with restriction fragment sequencing. PLoS One 8: e63960.

Suchan, T., C. Pitteloud, N. Gerasimova, A. Kostikova, N. Arrigo, M. Pajkovic, M. Ronikier, and N. Alvarez. 2015 Hybridization capture using RAD probes (hyRAD), a new tool for performing genomic analyses on museum collection specimens. BioRxiv doi:10.1101/025551.

Tin, M. M. Y., E. P. Economo, and A. S. Mikheyev. 2014. Sequencing degraded DNA from non-destructively sampled museum specimens for RAD-tagging and low-coverage shotgun phylogenetics. PLoS One 9: e96793.

Ward, L. D. and M. Kellis. 2012. Evidence of abundant purifying selection in humans for recently acquired regulatory functions. Science 337: 1675-1678.

Wagner, C. E., I. Keller, S. Wittwer, O. M. Selz, S. Mwaiko, L. Greuter, A. Sivasundar, and O. Seehausen. 2013. Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. Molecular Ecology 22: 787-798.

Wang, S., E. Meyer, J. K. McKay, and M. V. Matz. 2012. 2b-RAD: A simple and flexible method for genome-wide genotyping. Nature Methods 9: 808-810.

Wang, Z., M. Gerstein, and M. Snyder. 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics 10: 57-63.

Warren, W. C., D. F. Clayton, H. Ellegren, A. P. Arnold, L. W. Hillier, A. Künstner, S. Searle, S. White, A. J. Vilella, S. Fairley, A. Heger, L. Kong, C. P. Ponting, E. D. Jarvis, C. V. Mello, P. Minx, P. Lovell, T. A. F. Velho, M. Ferris, C. N.

Balakrishnan, S. Sinha, C. Blatti, S. E. London, Y. Li, Y.-C. Lin, J. George, J. Sweedler, B. Southey, P. Gunaratne, M. Watson, K. Nam, N. Backström, L. Smeds, B. Nabholz, Y. Itoh, O. Whitney, A. P. Pfenning, J. Howard, M. Völker, B. M. Skinner, D. K. Griffin, L. Ye, W. M. McLaren, P. Flicek, V. Quesada, G. Velasco, C. Lopez-Otin, X. S. Puente, T. Olender, D. Lancet, A. F. A. Smit, R. Hubley, M. K. Konkel, J. A.Walker, M. A. Batzer, W. Gu, D. D. Pollock, L. Chen, Z. Cheng, E. E. Eichler, J. Stapley, J. Slate, R. Ekblom, T. Birkhead, T. Burke, D. Burt, C. Scharff, I. Adam, H. Richard, M. Sultan, A. Soldatov, H. Lehrach, S. V. Edwards, S.-P. Yang, X. Li, T. Graves, L. Fulton, J. Nelson, A. Chinwalla, S. Hou, E. R. Mardis, R. K. Wilson. 2010. The genome of a songbird. Nature 464: 757–762.

Wetterstrand, M. S. 2015. DNA sequencing costs: Data from the NHGRI genome sequencing program (GSP). Available from http://www.genome.gov/sequencingcosts (last accessed September 23, 2015).

Willing, E. M., M. Hoffmann, J. D. Klein, D. Weigel, and C. Dreyer. 2011. Paired-end RAD-seq for de novo assembly and marker design without available reference. Bioinformatics 27: 2187-2193.

Yang, Z. and B. Rannala. 2010. Bayesian species delimitation using multilocus sequence data. Proceedings of the National Academy of Sciences of the U.S.A. 107: 9264-9269.

Zerbino, D. R., and E. Birney. 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Research 18: 821-829.

Zhang, G., B. Li, C. li, M. T. P. Gilbert, E. D. Jarvis, J. Wang, and The Avian Genome Consortium. 2014. Comparative genomic data of the Avian Phylogenomics Project. GigaScience 3: 26.

CHAPTER 3: GENOMIC VARIATION IN A WIDESPREAD NEOTROPICAL BIRD (*XENOPS MINUTUS*) REVEALS DIVERGENCE, POPULATION EXPANSION, AND GENE FLOW

INTRODUCTION

Lowland humid forests in the Neotropics contain some of the highest biodiversity on the planet (Pearson 1977). A number of hypotheses have been proposed to explain the origins of this diversity, most of which link biological diversification directly to tumultuous landscape changes that led to speciation via the geographic isolation of populations (Moritz et al. 2000; Antonelli et al. 2010). The hypotheses differ in the events and features implicated. These include the origins of major rivers in the Amazon basin (Sick 1967; Capparella 1987; Ribas et al. 2012), uplift of the Andes and other mountain ranges (Chapman 1917, 1926), past fragmentation of humid forest due to expansion of arid habitats (Haffer 1969) or marine transgressions (Nores 1999; Aleixo 2004), edaphic or climatic conditions associated with geologic arches (Lougheed et al. 1999; Wesselingh and Salo 2006), and areas of displacement due to invasion by temperate taxa during colder periods (Erwin 1979; Bush 1994).

Studies evaluating these hypotheses have typically addressed them using gene genealogies to infer the timing of divergence and the geographic location of vicariance. Using the conceptual framework of vicariance biogeography, researchers have searched for shared phylogeographic (or phylogenetic) relationships among taxa that would suggest a common mechanism of biological diversification (e.g., Cracraft and Prum 1988; Brumfield and Capparella 1996; Hall and Harvey 2002; Quijada-Mascareñas et al.

2007). In addition, molecular dating methods have been used to estimate the timing of population divergence events and to compare these dates to hypothesized events in the landscape evolution of the Neotropics (Patton et al. 2000; Weir 2006; Santos et al. 2009; Ribas et al. 2012). Although some general patterns have emerged from these studies, such as the importance of landscape features in delimiting populations and the absence of an increase in diversification during the Pleistocene, no single dominant model relating historical diversification to landscape history has emerged from decades of genetic studies (reviewed in Haffer 1997; Antonelli et al. 2010; Leite and Rogers 2013).

Interrogating processes beyond divergence may prove to be more fruitful in informing species histories (Takahata et al. 1995, Kuhner et al. 2009). For example, signatures of population size changes found in studies of Neotropical organisms (Aleixo 2004; Cheviron et al. 2005; Solomon et al. 2008; D'Horta et al. 2011) may evidence historical increases or decreases in habitat availability. Evidence of gene flow between populations, which may reveal instances of past connectivity between habitats or regions, has been uncovered in a few studies (Patton et al. 1994; Noonan and Gaucher 2005; Maldonado-Coelho et al. 2013). In addition, a few studies have detected the effects of natural selection and sexual selection among populations (Mallet 1993, Turner et al. 2004), which may be linked to past climatic changes or other events. Reconstructing how these diverse processes influenced modern phylogeographic patterns is challenging, but could provide new insight into the history of Neotropical diversification.

The availability of genome-scale datasets is improving inferences concerning the historical diversification of organisms (Li and Durbin 2011, Frantz et al. 2013). Genomic data, when combined with methods that account for coalescent stochasticity, allow for the

integration of information across many loci (Edwards and Beerli 2000), and provide greater statistical power for testing models of population history (Pool et al. 2010). Analyses based on genomic data result in narrower confidence intervals in estimates of important parameters such as divergence times, effective population sizes, and migration rates between populations (Smith et al. 2014). Dense sampling across the genome increases the probability of obtaining data from migrant alleles or genomic regions that have been influenced by selection (Carlson et al. 2005). The application of genomic data to Neotropical systems (e.g., Nadeau et al. 2013) promises to allow further investigation of processes important in Neotropical species histories.

Here, I examine the utility of dense, genome-scale genotyping-by-sequencing (GBS) data for phylogeography and historical demography. I use a GBS dataset from a widespread lowland Neotropical bird species (*Xenops minutus*; Aves, Furnariidae) to 1) characterize the geographic structure of genetic variation in this species and 2) evaluate a series of predictions concerning its historical demography. *Xenops minutus* is relatively common in humid lowland forests west of the Andes from Mexico to northwestern South America and, east of the Andes, in the Amazon Basin and Atlantic Forest of eastern South America (Remsen 2003). Eleven parapatrically or allopatrically distributed subspecies are currently recognized (Pinto 1954; Dickinson 2003; Remsen 2003). Subspecies are cryptic, varying subtly in plumage or vocalizations, but this variation has not been studied quantitatively. Remsen (2003) suggested that the nominate subspecies of southeastern Brazil is distinct in plumage and in its smaller size and may merit species status. Although all subspecies inhabit forest, it is unclear whether there is geographic variation in microhabitat preference or other ecological traits. Previous phylogeographic

studies (Burney 2009; Smith et al. 2014) of *X. minutus* had limited genomic or geographic sampling, but found evidence for geographically isolated mitochondrial clades and deep genome-wide divergence between populations from either side of the Andes, respectively. my goals were to determine how the population history of *X. minutus* influences modern patterns of genetic diversity, and to attempt to relate this history to the general landscape history of the Neotropics.

METHODS

Genetic Data Collection and Processing

We sampled eight vouchered *X. minutus* from each of nine biogeographic areas for a total of 72 individuals (Fig. 1, Appendix B). This sample included 7 of the 11 currently recognized subspecies (Dickinson 2003; Remsen 2003). The remaining four subspecies, distributed in Colombia, the northwestern Amazon Basin, and the northern Atlantic Forest of Brazil, were not included because I lacked sufficient genetic material. I extracted total DNA from frozen or alcohol-preserved pectoral muscle tissue using a DNeasy tissue extraction kit (Qiagen, Valencia, CA).

We sent 0.3-3.0 □g of each sample to the Cornell Institute of Genomic Diversity for genotyping-by-sequencing (GBS). GBS is a streamlined workflow for generating reduced representation libraries for Illumina sequencing, similar to other forms of RAD-Seq (Baird et al. 2008, Hohenlohe et al. 2010). Details of the laboratory methods can be

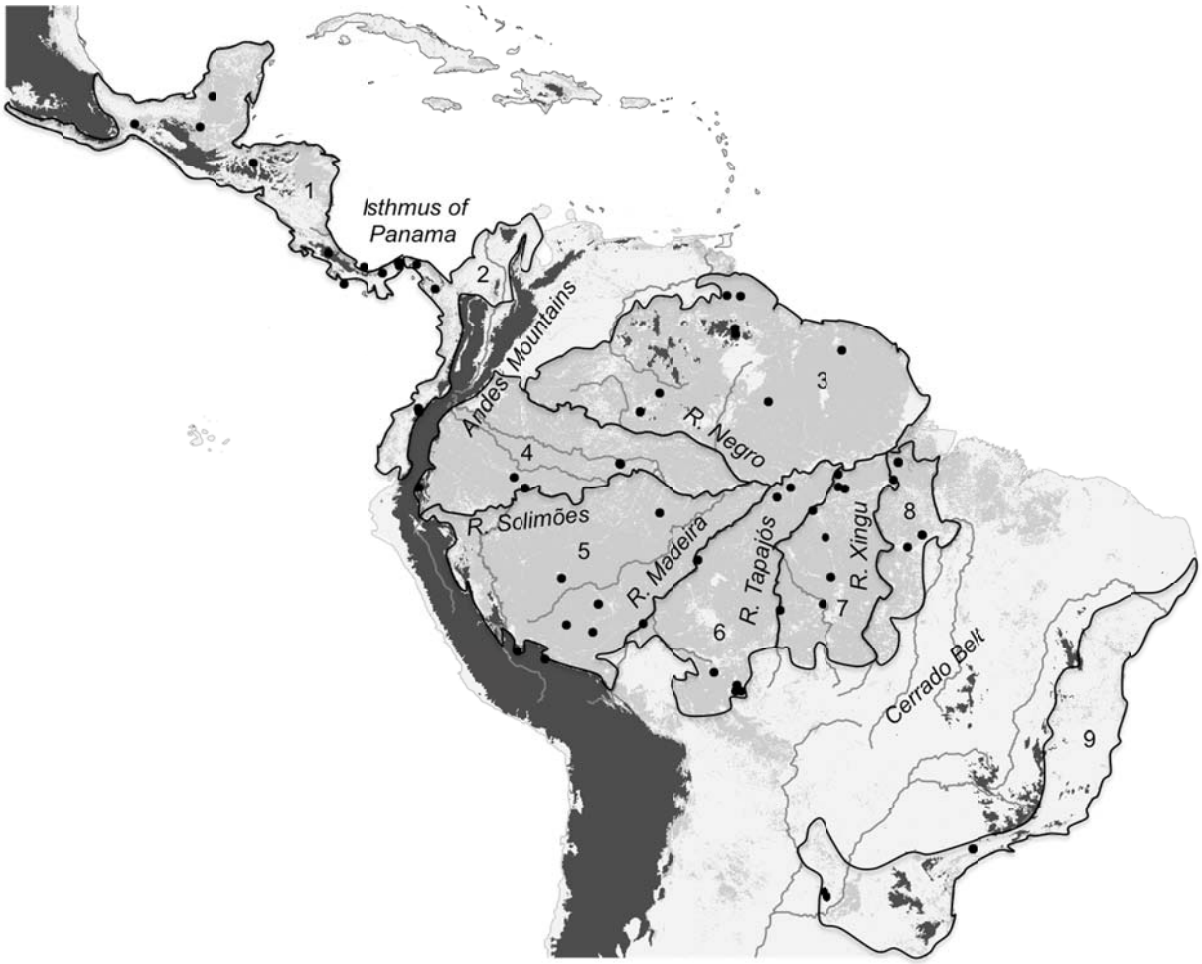


Figure 1. Map showing sampling locations (circles), biogeographic areas (bold type) and dispersal barriers (italics) examined in this study.

found in Elshire et al. (2011). In brief, DNA from each sample was digested using the restriction enzyme PstI (CTGCAG), and both a sample-specific barcoded adapter and a common adapter were ligated to the sticky ends of fragments. Samples were pooled and fragment libraries cleaned using a QIAquick PCR purification kit (Qiagen). Libraries were amplified using an 18-cycle PCR with long primers complementary to the barcoded and common adapters, purified again using QIAquick, and quantified using a PicoGreen assay (Molecular Probes, Carlsbad, CA, USA). Samples were run on a partial lane (72

out of 96 samples) of a 100-bp single-end Illumina HiSeq 2000 run at the Cornell Core Laboratories Center.

The Cornell Institute of Genomic Diversity processed raw sequence reads using the UNEAK pipeline, an extension to TASSEL 3.0 (Bradbury et al. 2007). Briefly, UNEAK retains all reads with a barcode, cut site, and no missing data in the first 64 bp after the barcode. Reads are clustered into tags by 100% identity, tags are aligned pairwise, and any tag pairs differing by one bp are called as potential SNPs. To remove sequencing errors, any alleles represented by fewer than five reads or a frequency of less than 5% are filtered out (Appendix B). Following processing with the UNEAK pipeline, I collapsed reverse complement tag-pairs and re-called genotypes using the method of Lynch (2009) as implemented in custom perl scripts obtained from T. A. White (White et al. 2013) and available at https://github.com/mgharvey/GBS_process_Tom_White/v1. I removed potential paralogs by filtering out SNPs with heterozygosity greater than 0.75, and I removed SNPs for which genotype calls were missing from more than 20% of the individuals. The hypothetical genomic distribution of the remaining SNP loci was investigated by aligning their tag-pair consensus sequences (with "N" inserted at the SNP site) to the Zebra Finch (*Taeniopygia guttata*) genome (Warren et al. 2010) using blastn (Altschul et al. 1990). *Taeniopygia guttata* is the most closely related species to *X. minutus* with a publicly available genome assembly, although the evolutionary distance between the two is considerable (Hackett et al. 2008). I used custom python scripts (available at http://github.com/mgharvey/misc_Python) to generate input files for further analysis.

Data analysis: Effects of distance and barriers

Isolation by distance and dispersal barriers are known to geographically structure genetic variation in Neotropical birds (Brawn et al. 1996; Cheviron et al. 2005; Cabanne et al. 2007). I evaluated the importance of these isolating forces using Mantel and partial Mantel tests, as well as a Bayesian model-based method. I used the kinship coefficient (Loiselle et al. 1995) calculated in the program SPAGeDi (Hardy and Vekemans 2002) as an index of pairwise genetic relatedness between individuals. The kinship coefficient F_{ij} is the probability that two homologous genes are identical by descent, and is calculated as $F_{ij} = (Q_{ij}-Q_m)/(1-Q_m)$ where Q_{ij} is the probability of identity by state between two individuals of interest for random genes and Q_m is the average probability of identity by state for genes coming from random individuals in the population. F_{ij} is a relatively unbiased estimator with low sampling variance (Hardy and Vekemans 2002).

We tested for isolation by distance across all individuals using a Mantel test comparing F_{ij} and geographic distance between individuals. Geographic distances were calculated as the Euclidean distances between sampling localities in SPAGeDi. To distinguish isolation by distance from discrete genetic breaks I conducted separate Mantel tests within each biogeographic area bounded by a major dispersal barrier, including the Isthmus of Panama, the Andes Mountains, major Amazonian rivers, and the cerrado belt of eastern Brazil that isolates Amazonia from the Atlantic Forest (based on Cracraft 1985, Fig. 1). To investigate isolation due to the dispersal barriers, I used a partial Mantel test that controlled for geographic distance in testing the correlation between F_{ij} and whether individuals were on the same or different sides of putative barriers. I conducted separate

analyses including all barriers and for each barrier individually. Only those individuals in the areas adjoining each barrier were used for the barrier-specific tests to remove confounding influences from other barriers. All Mantel and partial Mantel tests were carried out in the R package ecodist (Goslee and Urban 2007) using 10,000 permutations of geographic locations with individuals to determine significance and a jackknifing procedure to estimate standard errors.

Because Mantel and partial Mantel tests assume linear relationships between variables (Legendre and Fortin 2010), are confounded by spatial autocorrelation (Guillot and Rousset 2013), and are unable to directly quantify the relative importance of predictor variables (Bradburd 2013), I also used a new method, BEDASSLE (Bradburd 2013). BEDASSLE overcomes these issues by modeling the covariance in allele frequencies between populations as a function of the predictor variables, and estimating model parameters in a Bayesian framework using a Markov chain Monte Carlo algorithm. I used BEDASSLE to estimate the relative importance of geographic distance and barriers across the entire distribution of *X. minutus*, as well as between each pair of adjacent populations separated by a specific dispersal barrier. I ran BEDASSLE using the beta-binomial model to account for over-dispersion due to variation in demographic histories across populations. All analyses were run for 10 million generations, sampling every 100. I examined traces, marginal and joint marginal parameter distributions, and MCMC acceptance rates every one to five million generations and adjusted tuning parameters according to the suggestions of Bradburd et al. (2013).

Data analysis: Population assignment and admixture

We estimated the number of populations and conducted population assignment of individuals from all SNPs using methods implemented in STRUCTURE 2.3.4 (Pritchard et al. 2000) and Structurama (Huelsenbeck et al. 2011). Given a fixed number of populations (K), STRUCTURE assigns individuals to populations probabilistically such that Hardy-Weinberg equilibrium and linkage equilibrium within populations are maximized. In addition to population assignment, STRUCTURE can be used to identify admixed individuals. I used STRUCTURE without specifying prior information on population membership, and used options for correlated allele frequencies and genetic admixture across populations (Falush et al. 2003). I conducted runs of 1,000,000 generations (after a 10,000-generation burn-in) for each value between K=1 and K=15 and calculated Pr(X|K) to assess the results (Pritchard et al. 2000).

Structurama offers the option of jointly estimating the number of populations (K) and the assignment of individuals to populations using a Dirichlet process prior. I treated K as a random variable and provided an exponential distribution with a mean of nine as a prior for K, consistent with the number of biogeographic regions from which individuals were sampled. I also treated both K and the clustering variable α as random variables and examined the influence of three different gamma priors for α: (1,1), (5,1), and (10,1). For each analysis, I ran MCMC chains for 100 million generations, sampling every 25,000, and discarded 25% of the samples as burn-in.

To uncover finer scale population structure I used ChromoPainter and fineSTRUCTURE (Lawson et al. 2012) with the subset of SNPs having no missing data

across all 72 individuals. ChromoPainter considers each individual a possible recipient of "chunks" of DNA from a panel of donor individuals. It assembles a "coancestry matrix" recording the number of recombination events between each donor and recipient. In my case, I considered all individuals as potential recipients and donors. Although using linked sites provides more power for population inference using this method, I lacked linkage information for my SNPs, so I treated them as unlinked. fineSTRUCTURE then performs model-based clustering using the information in the coancestry matrix. The normalization parameter *c*, or the effective number of "chunks", is used to rescale the elements of the coancestry matrix before calculating the likelihood, and can influence the amount of inferred population structure. I used a *c* value of 1/(n-1) where n is the sample size, following the recommendation in Lawson et al. (2012) for unlinked data, but also examined the effects of higher and lower *c* values.

Population structure is sometimes inferred incorrectly due to the presence of isolation by distance (Meirmans 2012). I examined this possibility by conducting partial Mantel tests of the association between F_{ij} and both the set of populations estimated in fineSTRUCTURE and the set of populations estimated from STRUCTURE with K=5 and Structurama with the gamma prior for alpha equal to (1,5), while controlling for geographic distance. Hereafter I refer to these as the fineSTRUCTURE populations and the STRUCTURE/Structurama populations, respectively.

Data analysis: Population expansion and migration

We estimated expansion within and migration between both the fineSTRUCTURE and STRUCTURE/Structurama populations using coalescent modeling in the program LAMARC (Kuhner 2006, 2009). LAMARC has the advantage of being able to jointly estimate population growth and migration, both of which may be important processes influencing genetic variation in populations of tropical taxa (Moritz et al. 2000). I estimated the standardized population mutation rate ($\theta = 4N_e\mu$) and population growth rate (g, where $\theta_t = \theta_{present}^{-gt}$) for each population as well as the migration rate ($M = m/m\mu$, where m is the immigration rate per generation and $m\mu$ is the neutral mutation rate per site per generation) between adjacent populations separated by the dispersal barriers described above. I used the parameter-poor F84 model of sequence evolution because it is much faster than the alternative GTR model in LAMARC and because a simple model should be sufficient given that mutations are infrequent at the loci examined (SNPs represent a single variable site within an ~64 bp alignment). I set the transition/transversion ratio to 2. I used a Bayesian MCMC approach, and placed uniform priors on θ ($\log(1\times10^{-6}, 10)$), M ($\log(1\times10^{-10}, 100)$), and g (linear(-500, 1000)). I conducted 10 initial chains with 1,000 iterations of burn-in followed by 10,000 iterations, followed by 2 independent final chains of 5,000 iterations of burn-in followed by 10,000,000 iterations. I checked for convergence within and between chains using Tracer v.1.5 (Rambaut and Drummond 2007), and I report estimates from the second final chain.

Data analysis: Natural selection

We conducted a preliminary examination of selection in *X. minutus* using a multi-population outlier scanning approach implemented in BayeScan 2.01 (Foll and Gaggiotti 2008). BayeScan examines F_{st} values between each population and a common migrant gene pool for each locus. F_{st} coefficients are decomposed into a component shared by all loci (β) and a locus-specific component (α) that reflects selection. BayeScan then compares models in which selection (α) is and is not incorporated, and estimates the posterior probability for each model at each locus using a reversible-jump Markov chain Monte Carlo (RJ-MCMC) method. The posterior odds, or ratio of posterior probabilities, are used to decide on the best model and to define thresholds to determine sets of outlier markers. BayeScan is robust to complex demographic scenarios that might influence neutral differentiation (Foll and Gaggiotti 2008). I examined the influence of selection based on analyses using both the STRUCTURE/Structurama and fineSTRUCTURE populations. I ran analyses using 20 pilot runs of 5000 iterations, a burn-in of 50,000 iterations, and a final run of 50,000 iterations. Prior odds for the neutral model were set to 10.

Data analysis: Species tree

We estimated the branching structure of populations using a species tree approach for both the fineSTRUCTURE and STRUCTURE/Structurama populations. Species trees were estimated using the coalescent method implemented in SNAPP (Bryant et al. 2012).

SNAPP computes the likelihood of a species tree from unlinked biallelic markers rather than explicitly sampling gene trees. Any SNPs missing genotypes from all individuals in any of the populations were removed from the dataset. Also, due to the computational demands of analyzing the full dataset, I reduced each population to two randomly selected individuals (four haplotypes). I used a diffuse gamma prior for θ ($\alpha = 10$, $\beta = 100$) and a pure birth (Yule) prior for the species tree, with birth rate (λ) equal to 0.00765. For each population set, I conducted two runs of 5 million generations, sampling every 1,000 generations. I determined the burn-in and assessed MCMC convergence by examining ESS values and likelihood plots in Tracer v.1.5 (Rambaut and Drummond 2007). I combined runs and used TreeAnnotator (Rambaut and Drummond 2008) to determine the Maximum Clade Credibility tree and posterior probability values.

RESULTS

Sequencing and datasets

GBS produced a total of 106,784 biallelic SNPs (Appendix B). After collapsing reverse complements and filtering for observed heterozygosity and amount of missing data, the final data matrix contained 3,379 SNPs and was 91.1% complete. Data have been deposited in Dryad (submission pending). I recovered hits to the *T. guttata* genome using blastn for 3,247 of these SNPs. Hits were distributed across 31 of the 36 chromosomes, including the Z chromosome (Appendix B). The chromosomes without hits were microchromosomes 16, LGE22, LG2, LG5, and MT. The number of hits per

chromosome was positively associated with chromosome size ($r^2 = 0.836$, $p < 0.001$). I note, however, that the short length of GBS loci may result in low mapping accuracy and that *T. guttata* and *X. minutus* are distant relatives and synteny between the two genomes may be low.

Effect of distance and barriers on genetic divergence

Plotting pairwise kinship coefficients between samples relative to geographic distance revealed wide variation in kinship across the distribution of *Xenops minutus* (Fig. 2). Mantel tests showed a signal of isolation by distance based on correlations between the kinship coefficient and geographic distance (Mantel r statistic = -0.4964, p = 0.0001). However, the signal for isolation by distance was less prevalent within areas; only the Napo, Rondônia, and Atlantic Forest areas showed significant ($p < 0.01$) evidence of isolation by distance and the slopes were generally shallow (Appendix B). Partial Mantel tests across all areas and individuals revealed a relationship between kinship and whether individuals were on the same or opposite sides of barriers after controlling for isolation by distance (r = -0.6467, p = 0.0001). Examining each dispersal barrier separately, I found that all nine barriers showed a significant relationship ($p < 0.01$) with the kinship coefficient, and the slope of the Mantel correlation was generally steeper than in the within-area isolation by distance comparisons (Table 1, Appendix B). I observed the strongest correlations between dispersal barrier and kinship for the Isthmus of Panama, Andes Mountains, Rio Negro, and Rio Tapajós.

We discarded the first five million generations of all BEDASSLE MCMC chains and used the remaining posterior to estimate the ratio of the effect size of barriers versus the effect size of geographic distance (α_E/α_D). Across all barriers, the mean and median ratios were 0.413 and the 95% credible set was 0.322 to 0.464. The interpretation of this ratio is that the effect on genetic differentiation of separation by a barrier is equivalent to the effect of roughly 2,000 to 2,900 km of geographic distance. Examining each barrier

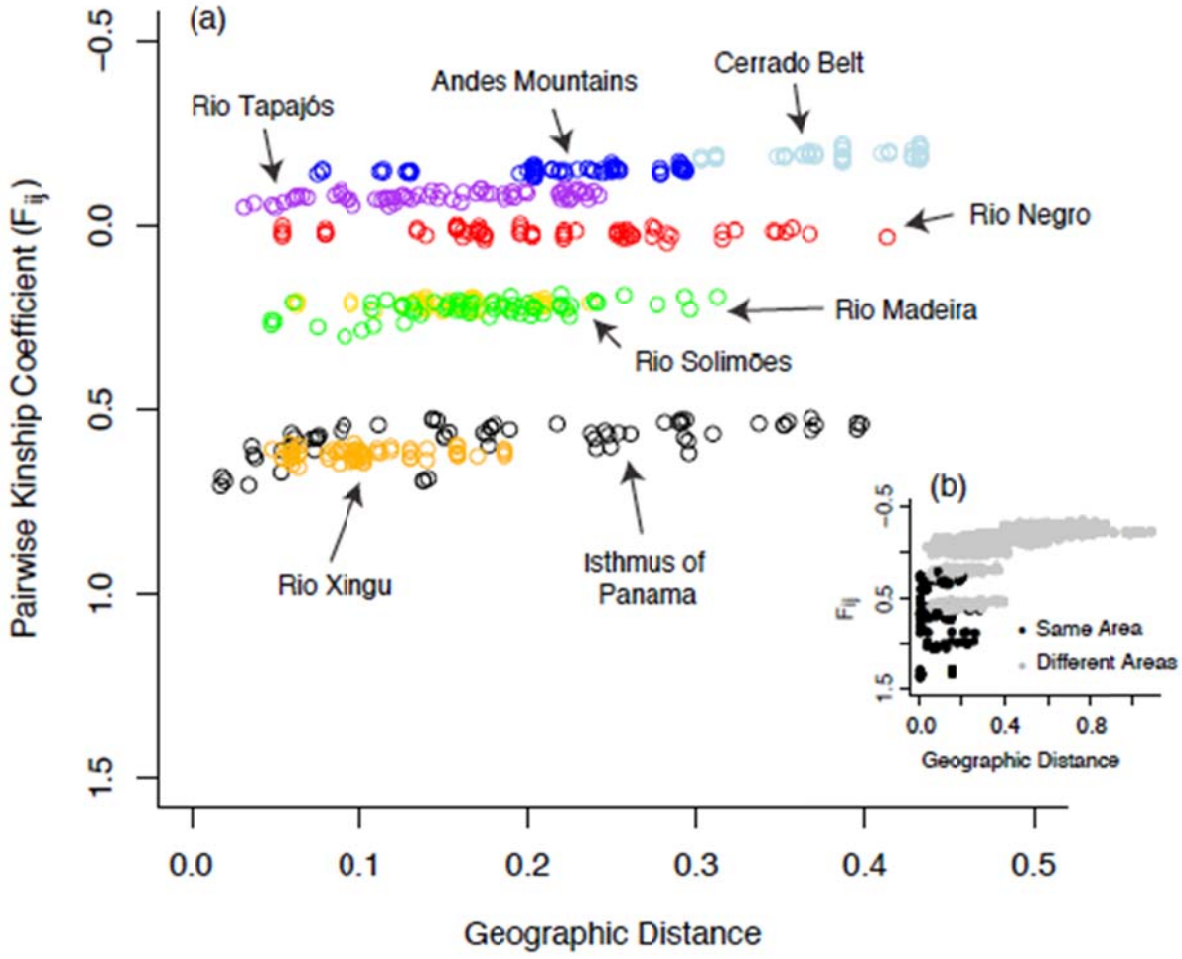


Figure 2. Plots of pairwise kinship versus relative geographic distance (a) between individuals separated by a single putative barrier and (b) between all individuals including those within the same area (black points) or separated by one or multiple barriers (gray points). The y-axes are inverted so that points representing greater divergence appear toward the tops of the plots.

separately, I found variation across barriers in the relative effect sizes of the barrier and geographic distance (Table 1). The Andes Mountains, Rio Negro, Rio Tapajós, and Cerrado Belt had the highest ratios, supporting the particular importance of these barriers in structuring genetic variation.

Table 1. Influence of barriers on genetic variation in *X. minutus*. Partial Mantel test r-statistics measure the relationship between pairwise kinship estimates and whether the two individuals are on the same or opposite sides of a barrier, controlling for geographic distance (lower r-statistics indicate a stronger relationship). The BEDASSLE α_E/α_D ratio measures the relative impact of barriers versus geographic distance on genetic similarity (higher values indicate a stronger relationship).

Dataset	partial Mantel test r-statistic (SE)	BEDASSLE αE/αD ratio (credible interval)
Isolation by Barriers		
All barriers	-0.647 (-0.676, -0.612)*	0.416 (0.276, 0.588)
Isthmus of Panama	-0.716 (-0.809, -0.646)*	0.0773 (0.0619, 0.0975)
Andes Mountains	-0.737 (-0.798, -0.620)*	137 (22.3, 466)
Rio Negro	-0.797 (-0.843, -0.736)*	62.2 (21.5, 129)
Rio Solimões	-0.519 (-0.830, -0.359)*	0.125 (0.0781, 0.189)
Rio Madeira	-0.469 (-0.661, -0.357)*	0.0168 (0.00905, 0.0271)
Rio Tapajós	-0.844 (-0.924, -0.800)*	99.0 (35.3, 324)
Rio Xingu	-0.276 (-0.410, -0.180)*	0.0296 (0.0150, 0.0682)
Cerrado Belt	-0.531 (-0.712, -0.421)*	136 (10.8, 8,060)

* *P<0.001*

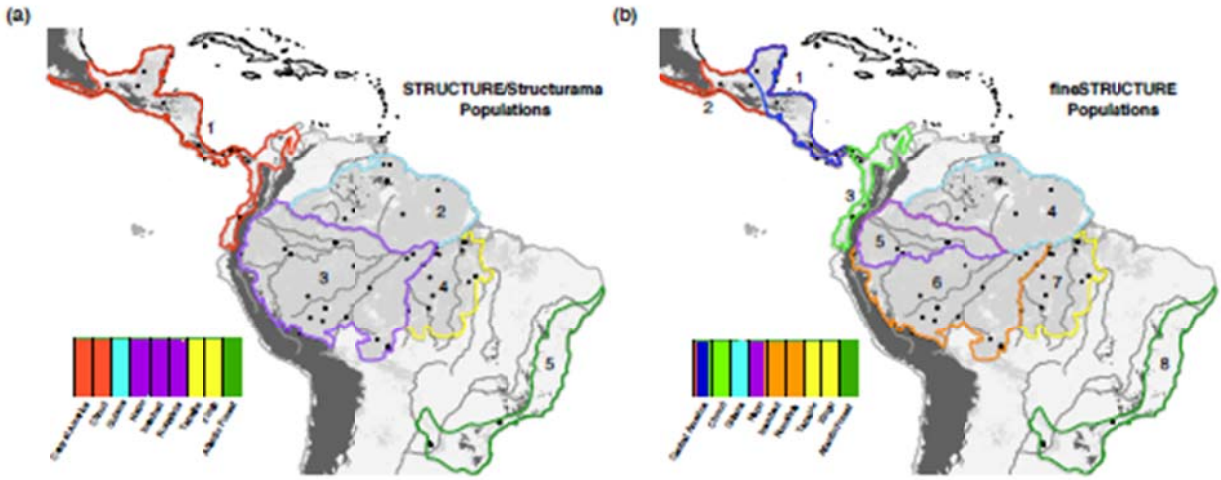


Figure 3. Maps of the distributions of populations from (a) the STRUCTURE/Structurama analysis and (b) the fineSTRUCTURE analysis. Populations are numbered and numbers are consistent with subsequent tables and figures. The adjacent structure plots show population membership for all individuals from (a) the STRUCTURE analysis with K=5 and (b) the fineSTRUCTURE analysis. Admixed individuals are shown in the structure plot for the STRUCTURE analysis, but fineSTRUCTURE does not estimate admixture.

Population assignment and admixture

Analysis of P(X|D) from the STRUCTURE runs suggested K=5 was the optimal value for number of populations (Appendix B). The five clusters from the K=5 analysis contained the individuals from (Central America + Chocó), Guiana, (Napo + Inambari + Rondônia), (Tapajós + Xingu), and Atlantic Forest (Fig. 3, Appendix B). The four populations from the K=4 analysis were similar, except the Guiana population was lumped with the (Napo + Inambari + Rondônia) population (Appendix B).

The Structurama results were sensitive to the specification of the α prior. The (1,1) prior, with a small mean and narrow s.d. resulted in three populations; the (5,1) prior with an intermediate mean and s.d. resulted in five populations; the (10,1) prior with a large mean and s.d. resulted in four populations; and the prior based on an expected

value of nine populations resulted in three populations (Appendix B). The populations from all analyses included some combination of the same populations from the STRUCTURE analysis. The five populations from the Structurama analysis with an intermediate prior of (5,1) were identical to the five populations from the STRUCTURE analysis at K=5 (Fig. 3, Appendix B). These five populations were selected for use in subsequent analyses.

fineSTRUCTURE revealed more population structure than did STRUCTURE and Structurama. For $c = 1/(n-1)$, eight populations were detected (Fig. 3, Appendix B). These resembled the populations from the STRUCTURE analysis with K=5 and the Structurama analysis with the (5,1) prior. fineSTRUCTURE, however, divided the (Central America + Chocó) population into two, with the break occurring west of the canal zone in Panama (an individual from Coclé just west of the canal is allied with the Chocó individuals), and identified a cluster within Central America comprising the two northwestern-most samples from foothill areas in Oaxaca and Chiapas, Mexico. In addition, fineSTRUCTURE separated seven of the eight individuals in the Napo region from those in the Inambari and Rondônia regions. The eighth sample from the Napo region allied with the Inambari and Rondônia samples. This sample was collected in the foothills of southern Ecuador not far from the Río Marañon, which is often considered the border between the Napo and Inambari regions. Varying the value of c within a narrow range did not strongly influence cluster assignment in fineSTRUCTURE, and did so in an intuitive manner (e.g. by combining two weakly divergent clusters). I selected the eight populations from the fineSTRUCTURE analysis with $c = 1/(n-1)$ for use in subsequent analyses.

Both the set of populations inferred from fineSTRUCTURE ($r = -0.6709$, $p = 0.0001$) and STRUCTURE/Structurama ($r = -0.7611$, $p = 0.0001$) explained kinship between individuals significantly, even after controlling for isolation by distance in partial Mantel tests (Table 1). An examination of the admixture estimates from the STRUCTURE analysis with K=5 revealed relatively low admixture between populations (Appendix B). A small amount of admixture was observed between Guiana and (Napo + Inambari + Rondônia) and between (Napo + Inambari + Rondônia) and (Tapajós + Xingu).

Population expansion and migration

LAMARC MCMC chains converged after 2-3 million generations, but were run to 20 million. In both the analyses of fineSTRUCTURE and STRUCTURE/Structurama populations, θ was smaller in the Atlantic Forest population than in all other populations except the Napo population in the fineSTRUCTURE analysis (Table 2). I detected significant population growth (confidence intervals not overlapping zero) in seven of the eight fineSTRUCTURE populations and all five of the STRUCTURE/Structurama populations (Table 2). Growth rates were higher in the (Tapajós + Xingu) and Atlantic Forest populations than in other populations, except for the Central American and Guianan populations in the analysis of fineSTRUCTURE populations.

We recovered significant non-zero migration rates (confidence intervals not overlapping zero) in six of the 14 pairwise estimates for the fineSTRUCTURE populations and three of the eight pairwise estimates for the STRUCTURE/Structurama

populations (Table 3). Migration between Central American and Mexican populations in the analysis of fineSTRUCTURE populations was higher than between most other populations. Migration was also detected from the Chocó region to Central America (fineSTRUCTURE), from the (Napo + Inambari + Rondônia) population to the trans-Andean populations (STRUCTURE/Structurama), and from the (Tapajós + Xingu) population to the Atlantic Forest (both analyses). Within the Amazon Basin, analysis of the STRUCTURE/Structurama populations detected migration in both directions across the Negro River, and analysis of the fineSTRUCTURE populations detected migration from the Napo to the Guianan and (Inambari + Rondônia) populations and from the (Inambari + Rondônia) population to the (Tapajós + Xingu) population.

Table 2. Theta (θ) and population growth rate (g) estimates from LAMARC for each STRUCTURE/Structurama and fineSTRUCTURE population (see Figure 3).

Population	θ (95% CI)	g (95% CI)
STRUCTURE/Structurama		
1	5.2 (2.9, 9.2)	64.4 (48.8, 75.3)
2	8.4 (2.2, 9.8)	70.6 (52.7, 94.3)
3	9.9 (6.9, 10.0)	55.7 (47.5, 63.1)
4	8.1 (3.7, 9.8)	120.6 (94.8, 133.8)
5	1.0 (0.4, 5.2)	174.3 (112.0, 241.3)
fineSTRUCTURE		
1	8.7 (0.4, 9.8)	91.9 (-170.2, 208.4)
2	5.7 (0.5, 9.5)	87.5 (57.7, 212.1)
3	5.2 (1.9, 9.5)	80.4 (54.5, 100.0)
4	9.5 (2.9, 9.9)	96.7 (68.2, 107.5)
5	2.6 (1.1, 5.7)	42.0 (32.7, 57.4)
6	9.9 (6.8, 10.0)	66.5 (57.0, 76.9)
7	8.1 (3.3, 9.8)	119.9 (90.7, 134.3)
8	1.1 (0.4, 3.9)	204.3 (120.6, 258.9)

Natural selection

We detected no loci putatively under diversifying selection using BayeScan with the STRUCTURE/Structurama populations and the false discovery rate (FDR) set to 0.05 (Appendix B). I did, however, detect 20 loci that were putatively under purifying or balancing selection (FDR=0.05). In the analysis of the fineSTRUCTURE populations I detected 32 loci putatively under diversifying selection and 41 loci putatively under purifying or balancing selection (FDR=0.05). Of the 20 loci putatively under

Table 3. LAMARC estimates of migration rate (M) between populations for both the STRUCTURE/Structurama populations and fineSTRUCTURE populations (see Figure 3).

Populations	M (95% CI)
STRUCTURE/Structurama	
1 -> 3	0.0 (0.0, 0.2)
3 -> 1	0.8 (0.0, 2.6)
2 -> 3	3.3 (0.9, 7.2)
3 -> 2	3.8 (0.4, 10.6)
3 -> 4	0.9 (0.0, 3.5)
4 -> 3	0.4 (0.0, 1.5)
4 -> 5	2.0 (0.1, 8.7)
5 -> 4	0.0 (0.0, 0.6)
fineSTRUCTURE	
1 -> 2	31.6 (2.5, 92.9)
2 -> 1	90.7 (12.5, 99.7)
1 -> 3	2.6 (0.0, 9.6)
3 -> 1	2.5 (0.1, 37.9)
3 -> 5	0.0 (0.0, 0.6)
5 -> 3	1.2 (0.0, 4.2)
4 -> 5	0.0 (0.0, 0.6)
5 -> 4	1.2 (0.0, 4.9)
5 -> 6	4.3 (2.0, 8.6)
6 -> 5	0.3 (0.0, 1.8)
6 -> 7	1.9 (0.2, 5.2)
7 -> 6	0.0 (0.0, 0.3)
7 -> 8	4.3 (0.1, 12.3)
8 -> 7	0.0 (0.0, 0.5)

purifying/balancing selection in the analysis of STRUCTURE/Structurama populations, 17 were also outliers putatively under purifying/balancing selection in the analysis of fineSTRUCTURE populations.

Species tree

We recovered well-supported topologies from the SNAPP species tree analyses of both the STRUCTURE/Structurama population set and the fineSTRUCTURE population set (PP of all nodes = 1.0). Runs converged after two to three million generations, so I used a burn-in of three million generations. I ran both runs for each set of populations an additional four million generations and used the combined sample of 4,000 trees to generate a Maximum Clade Credibility tree and posterior probability values for each node (Fig. 4). Topologies were consistent between the analysis of the STRUCTURE/Structurama populations and the analysis of the fineSTRUCTURE populations. Both estimated an initial divergence between the Atlantic Forest population and all other populations, followed by a divergence across the Andes. Within the Amazon Basin, both analyses estimated an earlier divergence across the Tapajós River followed by a subsequent divergence across the Negro River. Divergences between the two Central American populations, the Central American and Chocó populations, and the Napo and (Inambari + Rondônia) populations from the fineSTRUCTURE analysis were very shallow.

The SNP species tree was similar overall to a prior mitochondrial gene tree based on Cytochrome b data from the same samples used in this study (Smith et al. in review,

Fig. 4). It differed, however, in the placement of the Guianan population. In the SNP species trees, the Guianan population is sister to the (Napo + Inambari + Rondônia) clade with high support (PP = 1.0), and thus is nested within the clade containing the other Amazonian populations. In the mitochondrial gene tree, however, the Guianan population

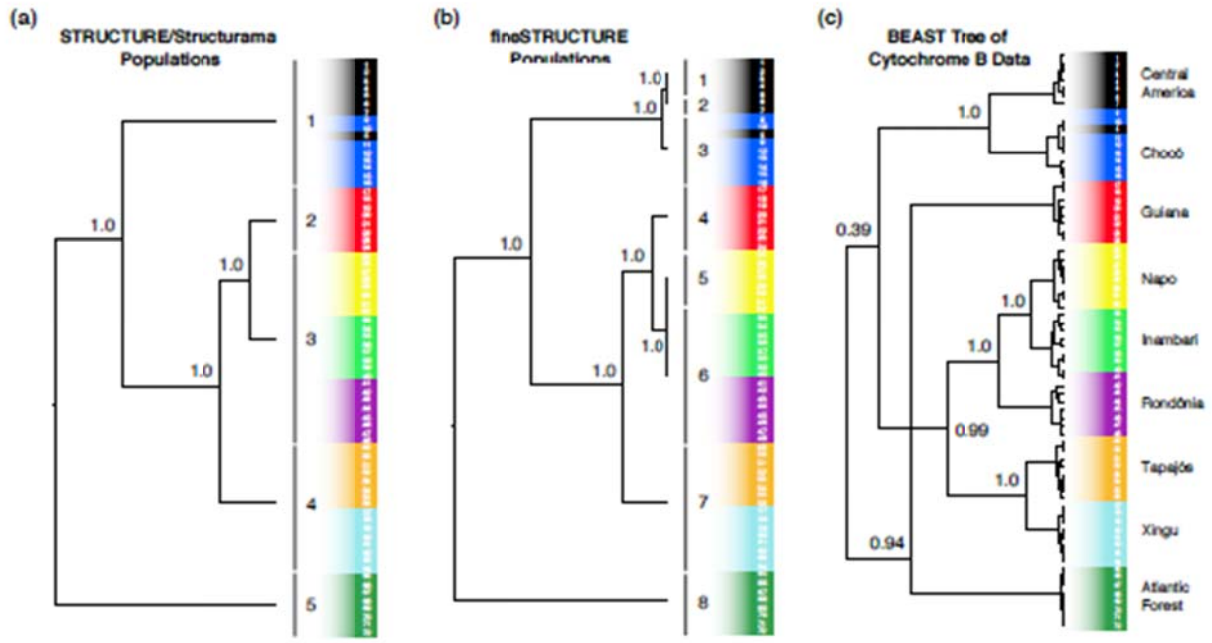


Figure 4 SNAPP species trees of (a) STRUCTURE/Structurama populations and (b) fineSTRUCTURE populations based on the SNP data and a (c) BEAST gene tree of sequence data from the mitochondrial gene Cytochrome B showing discordance with respect to the species trees. Colors are used to differentiate the areas of endemism from which individuals were sampled, and do not correspond to the population assignments from Figure 3.

is sister, albeit with a very long intervening branch, to the Atlantic Forest population with high support (PP = 0.94).

DISCUSSION

Prior studies of *Xenops minutus* based on mitochondrial sampling from many individuals (Burney 2009) or genomic sampling from a few individuals (Smith et al. 2014) revealed deep phylogeographic structure associated with major landscape features, such as the Andes mountains and Amazonian rivers. my GBS data identified the same phylogeographic breaks. Moreover, my results indicate the historical demography of *X. minutus* has been dynamic, with population size changes, migration and admixture between populations, and possibly natural selection.

We recovered positive population growth estimates for nearly all populations in the LAMARC analysis. Growth was greater in the (Tapajós + Xingu) and Atlantic Forest populations in the southeastern portion of the distribution than in most other populations. Signatures of population growth have been observed in some other Neotropical forest species (Aleixo 2004; Cheviron et al. 2005; Solomon et al. 2008; D'Horta et al. 2011, but see Lessa et al. 2003). The significant migration rates and evidence of admixture confirm that connectivity between currently isolated populations has occurred over the history of *X. minutus*. I recovered significant non-zero estimates for 9 of 22 total migration parameters across two different analyses in LAMARC. Across the Andes Mountains and cerrado belt, I detected significant migration in only one direction - out of rather than into the Amazon Basin. The STRUCTURE analysis also suggested the presence of limited admixture in some populations. In addition, I directly identified an admixed individual: the individual from the Napo region that clustered with the Inambari SNP clade. Prior mitochondrial data from this individual (Burney 2009) reveals a haplotype that clusters

closely with other Napo individuals, rather than individuals from the Inambari region (Fig. 4). This admixed individual therefore has a Napo mitochondrial haplotype, but an Inambari nuclear SNP genotype. There are few previous estimates of migration rate between populations of Neotropical forest organisms isolated by barriers, and these mostly suggest that gene flow is low or absent (Patton et al. 1994; Noonan and Gaucher 2005; Maldonado-Coelho et al. 2013). Hybridization and introgression between species and divergent forms have been uncovered in a few Neotropical taxa (Brumfield et al. 2001; Lovette 2004; Dasmahapatra et al. 2010; Naka et al. 2012). I expect that increased genomic representation in datasets will reveal that migration, hybridization, and introgression are an important part of the diversification history of the Neotropics.

Although I detected a small proportion of loci under purifying or balancing selection, the detection and interpretation of loci under purifying or balancing selection (ie. lower divergence than expected) is challenging (Teacher et al. 2013) due to the diversity of processes that might underlie such a pattern. The detection of diversifying selection at a small proportion of loci in the BayeScan analysis of fineSTRUCTURE populations, but not in the analysis of STRUCTURE/Structurama populations, suggested that diversifying selection has occurred between the most recently diverged populations. I found, however, that none of the outliers putatively under diversifying selection showed large allele frequency differences between populations that were only separated in the fineSTRUCTURE population set. Null F_{st} distributions may be overly narrow when some populations are recently diverged and have highly correlated allele frequencies, resulting in false positive outliers (Excoffier et al. 2009). Correlated allele frequencies between

recently diverged populations in the fineSTRUCTURE analysis, rather than diversifying selection, are likely responsible for the positive outliers in that analysis.

Accurately mapping loci to an annotated genome assembly may permit further evaluation of putative outliers (Stapley et al. 2010), but is complicated in my study by the absence of a genome assembly for *X. minutus* or any close relative, as well as the short length (~64 bp) of the GBS loci. Because I lack an independent method of verifying outliers, my results are very preliminary with regards to the importance of selection in this system. In addition to the problems mentioned above, the total number of loci putatively under selection across both BayeScan analyses (76 loci, 2.2% of the total) is smaller than in many other studies (reviewed in Nosil et al. 2009), suggesting a relatively minor role for selection in the history of *X. minutus*.

Relating species history to landscape history is challenging

Although I recovered a detailed estimate of the history of *X. minutus*, relating this history to the landscape history of the Neotropics and to hypotheses of Neotropical diversification in general is challenging. Similar issues have been encountered in other studies, such that few general patterns have emerged that convincingly relate landscape history to diversification history within species (Antonelli et al. 2010; Brumfield 2012). The difficulty stems in part from the incomplete knowledge of Neotropical landscape history on spatial and temporal scales relevant for species evolution (Bush 1994; Bush and Flenley 2007) and from the shortage of unique testable predictions under different hypotheses of Neotropical diversification (Brumfield and Capparella 1996; Tuomisto and

Ruokolainen 1997). Another challenge is that species distributions appear to be dynamic on much shorter timescales than those on which landscape evolution occurs, potentially erasing the signal for important events and resulting in pseudo-congruence (Haydon et al. 1994; Sanmartin et al. 2008; Brumfield 2012). Finally, different species are likely to have responded in different ways to the same history depending on their ecologies, such that few general patterns may exist (Aleixo 2006, Rull 2013).

We did find that major Neotropical landscape features, including the Andes, Amazonian rivers, and the cerrado belt isolating Amazonia from the Atlantic Forests, accounted for much of the genetic structure within *X. minutus*. The species tree topology for *X. minutus* contains similar area relationships to those found in other phylogenetic analyses (Weckstein and Fleischer 2005, Aleixo and Rossetti 2007). Divergence across barriers may be evidence of vicariance associated with barrier origin, dispersal across an existing barrier followed by differentiation (Mayr 1963), or the role of the barrier in structuring variation that arose elsewhere due to unknown historical processes (Brumfield 2012). The potential for pseudo-congruence between barriers and distributions combined with recent evidence that dispersal is more important than vicariance in the histories of some Neotropical groups (Fine et al. 2014, Smith et al. in review) suggests that the null hypothesis of shared area relationships used in vicariance biogeography is inappropriate. In addition, existing hypotheses of Neotropical diversification include few explicit predictions about relationships between areas of endemism (Bates et al. 1998, Leite and Rogers 2013), and replicate simulations illustrate a remarkable amount of phylogenetic discordance even under identical vicariance scenarios (Endler 1983). Because of these issues, the divergence patterns in *X. minutus* tell us relatively little about the historical

landscape or climatic events responsible for the modern genetic structuring in this species.

Dating the divergences between populations could allow determination of whether they were coincident with barrier formation, providing circumstantial support for particular vicariance hypotheses. Although dating the SNP divergences is problematic because I lack substitution rate estimates for GBS loci (see below), a previous dating analysis using mitochondrial DNA suggested that *X. minutus* populations diverged within the time span that the Andes Mountains and Amazonian Rivers are thought to have reached their modern conformations (Smith et al. in review). *Xenops minutus* populations across the Andes diverged 4.58 (s.d.=3.04-5.98) Mya and populations within the Amazon basin (aside from the Guianan population with a potential spurious placement in the mitochondrial tree, see below) began diverging 2.91 (s.d.=1.89-4.00) Mya. Similar Pliocene divergence dates have been estimated for many other Neotropical taxa including fish (e.g., Lovejoy et al. 2010; Lundberg et al. 2010), plants (e.g., Pennington and Dick 2010), amphibians (e.g., Santos et al. 2009), birds (e.g., Weir 2006), and mammals (e.g., Costa 2003). These dates coincide roughly with the final uplift of the Andes and the coincident formation of the contemporary fluvial system of the Amazon in the last 10 My (Mora et al. 2010). However, the concordance of divergence dates with the vast time span associated with the origin of these dispersal barriers provides only rough, circumstantial support. The crucial details of how dispersal barriers interdigitate with other factors, such as population size flux, changes in forest distribution (Bush and Flenley 2007), changes in forest composition and niche availability (Jaramillo et al. 2010), changes in avian community composition (Ricardo Negri et al. 2010), and local extinctions and re-

colonizations are not considered. This uncertainty suggests a nuanced understanding of how the Andes and Amazonian rivers influence speciation within lineages is not achievable using area relationships and divergence dates, and that my focus should be on other aspects of the speciation process.

The evidence I found for population expansions in *X. minutus* provides support for a prediction of the forest refugia hypothesis that humid lowland forests were once more restricted due to the expansion of savanna (Haffer 1969). Some palynological analyses also support the idea that lowland Neotropical humid forest was once more restricted (Absy et al. 1991; Burnham and Graham 1999). Recent isotopic evidence suggests that precipitation was lower in the eastern Amazon, but not the western Amazon, during the last glaciation (Cheng et al. 2013), consistent with my observation of greater population growth in that area. Unfortunately, knowledge of the recent history of forest cover in the Amazon is limited and contentious (Behling et al. 2010). The marine incursion hypotheses might also predict population growth following the recession of water levels, although growth is expected to be greatest in the western Amazon Basin (Aleixo 2004), contrary to the pattern I observed. Other events such as disease (e.g., Daszak et al. 2003), changes in abiotic climate conditions (e.g., Sillett et al. 2000), or changes in competitive interactions (e.g., Koenig 2003), predation (e.g., Wittmer et al. 2005), or resource availability (e.g., O'Donoghue et al. 1997) might also have driven population size changes. Although the population expansion I observed in *X. minutus* may be attributable to recent increases in forest habitat in the lowland Neotropics, I cannot exclude other equally likely causes.

Migration and admixture between populations supports the idea that populations have experienced periodic connections in the past. Habitat connectivity, however, might have occurred under any of various hypotheses of Neotropical diversification and does not aid in discriminating among them. Future improvements in my understanding of past habitat distributions combined with improved methods of inferring and dating admixture events may allow us to correlate episodes of migration and gene flow with individual events of habitat connectivity (Gillespie and Roderick 2014).

Based on the challenges associated with connecting the species history of *X. minutus* to landscape history, I suggest the common practice of relating single species histories to landscape events is unproductive. As an alternative, I suggest an initial focus on evaluating the importance of different historical processes (including divergence, but also population size changes, migration, and selection) using genomic datasets within individual species or species complexes. With many such datasets in hand, comparative methods may permit determination of the importance of each process along taxonomic, temporal, and spatial axes. This information, perhaps combined with more information on the combined effects of processes shaping landscape history, may ultimately permit evaluation of each hypothesis of Neotropical diversification across assemblages, timescales, and regions.

Limitations and prospects for GBS data in phylogeography

Genotyping-by-sequencing data allowed us to conduct a variety of population genetic, phylogeographic, and phylogenetic analyses. I did, however, encounter some

potential shortcomings of GBS data for addressing phylogeographic questions in my non-model system. The large amount of missing data observed in my dataset prior to filtering suggests the need for further optimization of coverage relative to the number of targeted loci, but better coverage could be achieved by using different enzymes or multiple enzymes (Peterson et al. 2012). The locations to which I were able to map loci may be inaccurate, both because of the potential for spurious alignment due to the short length of the GBS reads, and because of the evolutionary distance between *X. minutus* and *T. guttata*. This issue may be reduced in the future if longer read lengths can be obtained, or if a genome from a species more closely related to the study species becomes available. Perhaps the greatest limitation of GBS is that no standard evolutionary rate exists for the targeted loci for the purpose of dating divergences or converting demographic parameters. As a result, I were largely limited to making relative comparisons of raw parameter estimates in this study. Furthermore, the processing pipeline for GBS and other RAD-Seq data complicates the future development of standard rates that could be used across groups of organisms. Because identity thresholds are applied to each dataset for assembly, datasets may be truncated to different degrees and rates are not directly comparable. More informed assembly protocols or methods for correcting rates based on the level of truncation in a dataset may alleviate these issues in the future.

Despite some limitations, genomic data from GBS have provided a more complete picture of the history of *X. minutus* than would be possible with a few markers. The history inferred from genomic SNPs is likely to better reflect the true history of *X. minutus* populations than a single-locus dataset (Edwards and Beerli 2000). In addition, genomic data have allowed us to investigate processes that are difficult to evaluate with a

single marker, such as migration and selection. More efficient laboratory methods and new analytical tools will surely increase the utility of genomic datasets as they come into more widespread use.

Since divergence histories based on mitochondrial data have been the primary source of information for studies of Neotropical phylogeography (Haffer 1997; Antonelli et al. 2010; Leite and Rogers 2013), the discordance between the mitochondrial gene tree and genome-wide SNP species trees in this study is alarming. This discrepancy might occur if deep coalescence of the mitochondrial haplotypes from the Guianan and Atlantic Forest populations resulted in a mitochondrial genealogy that does not represent the species history. Alternatively, recent nuclear gene flow between Atlantic Forest and Guianan populations might produce a similar result, but I consider this less likely due to the geographic distance between these populations and because gene flow would have to have influenced a substantial portion of the genome to result in the relationship recovered from the GBS loci. Discordance between mitochondrial and nuclear SNP datasets is not surprising, given the number of prior studies reporting similar mito-nuclear discordance (Funk and Omland 2003; Chan and Levin 2005). The observed discordance deepens concerns about the utility of mitochondrial DNA as a record of population history and reaffirms the importance of shifting to genome-wide datasets for phylogeographic research.

Systematics of Xenops minutus

Our results support the presence of at least three deeply divergent clades experiencing little to no gene flow within *Xenops minutus*. The trans-Andean clade of Central and northwestern South America includes the subspecies *mexicanus* (Sclater, 1857), *ridgwayi* (Hartert and Goodson, 1917), and *littoralis* (Sclater, 1862). The trans-Andean subspecies *olivaceus* Aveledo and Pons, 1952 and *neglectus* (Todd, 1913) were not sampled in my study or previous studies but may also below to this group. The Amazonian/Guianan clade includes the subspecies *genibarbis* (Illiger, 1811); *obsoletus* Zimmer, 1924; and *ruficaudus* (Vieillot, 1816). The northwestern Amazonian subspecies *remoratus* Zimmer, 1935, not sampled in my study, may also be in this group, although mitochondrial data suggest that this population is highly divergent (Burney 2009). Populations from the northern Atlantic Forest are most similar to the Amazonian/Guianan clade based on mitochondrial data (Burney 2009). These populations were described as a unique subspecies (*alagoanus* Pinto, 1954), but this taxon has been omitted or overlooked by most subsequent authors (Dickinson 2003; Remsen 2003) and was not sampled in my study. Finally, the nominate subspecies (Sparrman, 1788) of the Atlantic Forest represents the third deeply divergent clade, and is highly distinct genetically despite some amount of gene flow from Amazonian populations to the northwest.

All three clades are diagnosable vocally and some show plumage differences. The trans-Andean clade has a much more rapid, nearly trilled, song than other clades. The Amazonian/Guianan clade has a slower song with rising, “hill-shaped” (Isler et al. 1998) notes. The nominate subspecies also has a slow song, but the notes are upslurred giving

them a distinct "twanging" quality. Interestingly, populations from the northern Atlantic Forest (*alagoanus*) have a song more similar to Amazonian birds, and thus may be part of the Amazonian/Guianan clade. Plumage is variable geographically, but much of the variation appears to be clinal (Remsen 2003). Within the trans-Andean clade, plumage is highly variable with a rough trend from red and plain in the north to olive and streaked in the south. Plumage is also variable in the Amazonian/Guianan clade, although most populations are intermediate in color and show moderate to heavy streaking. Only the nominate subspecies is highly distinct in plumage (Remsen 2003), with a white throat, reddish coloration, and plain underparts.

We suggest that the three deeply divergent clades described above represent phylogenetic species due to diagnosable vocal, genetic, and (in the third clade) plumage differences. They may merit biological species status based on the fact that they exhibit little to no detectable gene flow, although further research is required to determine whether they might currently interbreed. The northwestern Amazonian clade found by Burney (2009) may represent a fourth phylogenetic species, although it would be desirable to confirm this result with additional independent genetic markers, vocal data, and field work to determine if populations come into contact in the northwestern Amazon Basin. The populations from Guiana and from the Tapajós/Xingu areas of endemism may also merit species status because they were recovered as distinct populations and show moderate divergence in the species tree. These two clades are less divergent, however, than the three mentioned above, and I were also unable to find obvious morphological or vocal characters distinguishing them. Further research involving improved geographic

sampling and formal morphological and vocal analyses may clarify the status of these and other, un-sampled populations.

REFERENCES

Absy, M., A. Cleef, M. Fournier, L. Martin, M. Servant, A. Sifeddine, M. Ferreira da Silva, F. Soubies, K. Suguio, B. Turcq, and T. van der Hammen. 1991. Mise en évidence de quatre phases d'ouverture de la forêt dense dans le sud-est de l'Amazonie au cours des 60,000 dernières années. Première comparaison avec d'autres régions tropicales. Comptes rendus de l'Académie des Sciences de Paris 312: 673-678.

Aleixo, A. 2004. Historical diversification of a terra-firme forest bird superspecies: a phylogeographic perspective on the role of different hypotheses of Amazonian diversification. Evolution 58: 1303-1317.

Aleixo, A. 2006. Historical diversification of floodplain forest specialist species in the Amazon: a case study with two species of the avian genus *Xiphorhynchus* (Aves: Dendrocolaptidae). Biological Journal of the Linnean Society 89: 383-395.

Aleixo, A. and D. F. Rossetti. 2007. Avian gene trees, landscape evolution, and geology: towards a modern synthesis of Amazonian historical biogeography? Journal of Ornithology 148: 443-453.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. Journal of Molecular Biology 215: 403-410.

Antonelli, A., A. Quijada-Mascareñas, A. J. Crawford, J. M. Bates, P. M. Velasco, W. Wüster. 2010. Molecular studies and phylogeography of Amazonian tetrapods and their relation to geological and climatic models. Pages 386-404 in Amazonia, Landscape and Species Evolution: A Look into the Past (C. Hoorn and F. P. Wesselingh, eds.). Wiley-Blackwell, London, UK.

Aveledo, R. and A. R. Pons. 1952. Aves nuevas y extensiones de distribución a Venezuela. Novedades Cientificas - Contribuciones Ocasionales del Museo de Historia Natural La Salle 7: 7-8.

Baird, N. A., P. D. Etter, T. S. Atwood, M. C. Currey, A. L. Shiver, Z. A. Lewis, E. U. Selker, W. A. Cresko, and E. A. Johnson. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One 3: e3376.

Bates, J. A., S. J. Hackett, J. Cracraft. 1998. Area-relationships in the Neotropical lowlands: an hypothesis based on raw distributions of passerine birds. Journal of Biogeography 25, 783-793.

Behling, H., M. Bush, and H. Hooghiemstra. 2010. Biotic development of quaternary Amazonia: a palynological perspective. Pages 335-345 in Amazonia, Landscape and Species Evolution: A Look into the Past (C. Hoorn and F. P. Wesselingh, eds.). Wiley-Blackwell, London, UK.

Bradburd, G. S., P. L. Ralph, and G. M. Coop. 2013. Disentangling the effects of geographic and ecological isolation on genetic differentiation. Evolution 67: 3258-3273.

Bradbury, P. J., Z. Zhang, D. E. Kroon, T. M. Casstevens, Y. Ramdoss, and E. S. Buckler. 2007. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23: 2633-2635.

Brawn, J. D., T. Collins, M. Medina, and E. Bermingham. 1996. Associations between physical isolation and geographical variation within three species of Neotropical birds. Molecular Ecology 5: 33-46.

Brumfield, R. T. 2012. Inferring the origins of lowland Neotropical birds. Auk 129: 367-376.

Brumfield, R. T. and A. P. Capparella. 1996. Historical diversification of birds in northwestern South America: a molecular perspective on the role of vicariant events. Evolution 50: 1607-1624.

Brumfield, R. T., R. W. Jernigan, D. B. McDonald, and M. J. Braun. 2001. Evolutionary implications of divergent clines in an avian (*Manacus*: Aves) hybrid zone. Evolution 55: 2070-2087.

Bryant, D., R. Bouckaert, J. Felsenstein, N. A. Rosenberg, and A. RoyChoudhury. 2012. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. Molecular Biology and Evolution 29: 1917-1932.

Burney, C. W. 2009. Comparative phylogeography of Neotropical birds. PhD Dissertation. Louisiana State University, Baton Rouge, LA.

Burnham, R. J. and A. Graham. 1999. The history of neotropical vegetation: new developments and status. Annals of the Missouri Botanical Garden 86: 546-589.

Bush, M. B. 1994. Amazonian speciation: a necessarily complex model. Journal of Biogeography 21: 5-17.

Bush, M. B. and J. Flenley. 2007. Tropical Rainforest Responses to Climatic Change. Springer, New York, USA.

Cabanne, G. S., F. R. Santos, and C. Y. Miyaki. 2007. Phylogeography of *Xiphorhynchus fuscus* (Passeriformes, Dendrocolaptidae): vicariance and recent demographic expansion in southern Atlantic forest. Biological Journal of the Linnaean Society 91: 73-84.

Capparella, A. P. 1987. Effects of riverine barriers on genetic differentiation of Amazonian forest undergrowth birds. PhD Dissertation. Louisiana State University, Baton Rouge, LA.

Carlson, C. S., Thomas, D.J., Eberle, M.A., Swanson, J.E., Livingston, R.J., Reider, M.J., Nickerson, D.A., 2005. Genomic regions exhibiting positive selection identified from dense genotype data. Genome Research 15: 1553-1565.

Chan, K. and S. A. Levin. 2005. Leaky prezygotic isolation and porous genomes: rapid introgression of maternally inherited DNA. Evolution 59: 720-729.

Chapman, F. M. 1917. The distribution of bird-life in Colombia: a contribution to a biological survey of South America. Bulletin of the American Museum of Natural History 36: 1-729.

Chapman, F. M. 1926. The distribution of bird-life in Ecuador. Bulletin of the American Museum of Natural History 55: 1-784.

Cheng, H., A. Sinha, F. W. Cruz, X. Wang, R. Lawrence Edwards, F. M. d'Horta, C. C. Ribas, M. Vuille, L. D. Stott, A. S. Auler. 2013. Climate change patterns in Amazonia and biodiversity. Nature Communications 4: 1411.

Cheviron, Z., S. J. Hackett, and A. P. Capparella. 2005. Complex evolutionary history of a Neotropical lowland forest bird (*Lepidothrix coronata*) and its implications for historical hypotheses of the origin of Neotropical avian diversity. Molecular Phylogenetics and Evolution 36: 338-357.

Costa, L. P. 2003. The historical bridge between the Amazon and the Atlantic Forest of Brazil: a study of molecular phylogeography with small mammals. Journal of Biogeography 30: 71-86.

Cracraft, J. 1985. Historical biogeography and patterns of differentiation within the South American avifauna: areas of endemism. Ornithological Monographs 36: 49-84.

Cracraft, J., and R. O. Prum. 1988. Patterns and processes of diversification: speciation and historical congruence in some Neotropical birds. Evolution 42: 603-620.

D'Horta, F. M., G. S. Cabanne, D. Meyer, and C. Y. Miyaki. 2011. The genetic effects of Late Quaternary climatic changes over a tropical latitudinal gradient: diversification of an Atlantic Forest passerine. Molecular Ecology 20: 1923-1935.

Dasmahapatra, K. K., G. Lamas, F. Simpson, and J. Mallet. 2010. The anatomy of a ''suture zone'' in Amazonian butterflies: a coalescent-based test for vicariant geographic divergence and speciation. Molecular Ecology 19: 4283-4301.

Daszak, P., A. A. Cunningham, and A. D. Hyatt. 2003. Infectious disease and amphibian population declines. Diversity and Distributions 9: 141-150.

Dickinson, E. C. 2003. The Howard and Moore Complete Checklist of the Birds of the World. Christopher Helm, London, UK.

Edwards, S. and P. Beerli. 2000. Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. Evolution 54: 1839-1854.

Elshire, R. J., J. C. Glaubitz, Q. Sun, J. A. Poland, K. Kawamoto, E. S. Buckler, and S. E. Mitchell. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6: e19379.

Endler, J. A. 1983. Testing causal hypotheses in the study of geographical variation. Pages 424-443 in Numerical Taxonomy (J. Felsenstein, ed.). Springer-Verlag, Berlin, Germany.

Erwin, T. L. 1979. Thoughts on the evolutionary history of ground beetles: hypotheses generated from comparative faunal analyses of lowland forest sites in temperate and tropical regions. Pages 539-592 in Carabid Beetles, Their Evolution, Natural History, and Classification (T. L. Erwin, G. E. Ball, D. R. Whitehead, and A. L. Halpern, eds.). DW Junk, The Hague, Netherlands.

Excoffier, L., T. Hofer, and M. Foll. 2009. Detecting loci under selection in a hierarchically structured population. Heredity 103: 285-298.

Falush, D., M. Stephens, and J. K. Pritchard. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164: 1567-1587.

Fine, P.V.A., F. Zapata, and D. C. Daly. 2014. Investigating processes of Neotropical rain forest tree diversification by examining the evolution and historical biogeography of the Protieae (Burseraceae). Evolution 68: 1988-2004.

Foll, M. and O. Gaggiotti. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics 180: 977-993.

Frantz, L. A. F., J. G. Schraiber, O. Madsen, H.-J. Megens, M. Bosse, Y. Paudel, G. Semiadi, E. Meijaard, N. Li, R. P. M. A. Crooijmans, A. L. Archibald, M. Slatkin, L. B. Schook, G. Larson, and M. A. M. Groenen. 2013. Genome sequencing reveals fine scale diversification and reticulation during speciation in *Sus*. Genome Biology 14: R107.

Funk, D. J. and K. E. Omland. 2003. Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. Annual Review of Ecology, Evolution, and Systematics 34: 397-423.

Gillespie, R. G. and G. K. Roderick. 2014. Evolution: geology and climate drive diversification. Nature 509: 297-298.

Guillot, G. and F. Rousset. 2013. Dismantling the mantel tests. Methods in Ecology and Evolution 4: 336-344.

Goslee, S. C. and D. L. Urban. 2007. The ecodist package for dissimilarity-based analysis of ecological data. Journal of Statistical Software 22: 1-19.

Hackett, S. J., R. T. Kimball, S. Reddy, R. C. K. Bowie, E. L. Braun, M. J. Braun, J. L. Chojnowski, W. A. Cox, K.-L. Han, J. Harshman, C. J. Huddleston, B. D. Marks, K. J. Miglia, W. S. Moore, F. H. Sheldon, D. W. Steadman, C. C. Witt, and T. Yuri. 2008. A phylogenomic study of birds reveals their evolutionary history. Science 320: 1763-1768.

Haffer, J. 1969. Speciation in Amazonian forest birds. Science 165: 131-137.

Haffer, J. 1997. Alternative models of vertebrate speciation in Amazonia: an overview. Biodiversity Conservation 6: 451-476.

Hall, J.P. and D. J. Harvey. 2002. The phylogeography of Amazonia revisited: new evidence from riodinid butterflies. Evolution 56: 1489-1497.

Hardy, O.J. and X. Vekemans. 2002. SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Molecular Ecology Notes 2: 618-620.

Hartert, E. and F. W. Goodson. 1917. Notes and descriptions of South American birds. Novitates Zoologicae 24: 410-419.

Haydon, D. T., R. R. Radtkey, and E. R. Pianka. 1994. Experimental biogeography: interactions between stochastic, historical, and ecological processes in a model archipelago. Pages 117-130 in Species Diversity in Ecological Communities: Historical and Geographical Perspectives (R. E. Ricklefs and D. Schluter, eds.). University of Chicago Press, Chicago, USA.

Hohenlohe, P. A., S. Bassham, P. D. Etter, N. Stiffler, E. A. Johnson, and W. A. Cresko. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genetics 6: e1000862.

Huang, H. and L. L. Knowles. 2014. Unforeseen consequences of excluding missing data from next-generation sequences: Simulation study of RAD sequences. Systematic Biology doi:10.1093/sysbio/syu046.

Huelsenbeck, J.P., P. Andolfatto, and E. T. Huelsenbeck. 2011. Structurama: Bayesian inference of population structure. Evolutionary Bioinformatics Online 7: 55-59.

Illiger, D., 1811. C. Prodromus Systematis Mammalium et Avium. C. Salfeld, Berlin, Germany.

Ilut, D. C., M. L. Nydam, and M. P. Hare. 2014. Defining loci in restriction-based reduced representation genomic data from nonmodel species: Sources of bias and diagnostics for optimal clustering. BioMed Research International 2014: 675158.

Isler, M.L., P. R. Isler, and B. M. Whitney. 1998. Use of vocalizations to establish species limits in antbirds (Passeriformes: Thamnophilidae). Auk 115: 577-590.

Jaramillo, C., C. Hoorn, S. A. F. Silva, F. Leite, F. Herrera, L. Quiroz, R. Dino, and L. Antonioli. 2010. The origin of the modern Amazon rainforest: implications of the palynological and palaeobotanical record. Pages 317-334 in Amazonia, Landscape and Species Evolution: A Look into the Past (C. Hoorn and F. P. Wesselingh, eds.). Wiley-Blackwell, London, UK.

Koenig, W. D. 2003. European starlings and their effect on native cavity-nesting birds. Conservation Biology 17: 1134-1140.

Kuhner, M. K. 2006. LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. Bioinformatics 22: 768-770.

Kuhner, M. K., 2009. Coalescent genealogy samplers: windows into population history. Trends in Ecology and Evolution 24: 86-93.

Lawson, D. J., G. Hellenthal, S. Myers, and D. Falush, D., 2012. Inference of population structure using dense haplotype data. PLoS Genet. 8: e1002453.

Legendre, P. and M. J. Fortin. 2010. Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. Molecular Ecology Resources 10: 831-844.

Leite, R. N. and D. S. Rogers. 2013. Revisiting Amazonian phylogeography: insights into diversification hypotheses and novel perspectives. Organismal Diversity and Evolution 13: 639-664.

Lessa, E.P., J. A. Cook, and J. L. Patton. 2003. Genetic footprints of demographic expansion in North America, but not Amazonia, during the Late Quaternary. Proceedings of the National Academy of Sciences of the U.S.A. 100: 10331-10334.

Li, H. and R. Durbin. 2011. Inference of human population history from individual whole-genome sequences. Nature 475: 493-496.

Loiselle, B.A., V. L. Sork, J. Nason, and C. Graham. 1995. Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). American Journal of Botany 82: 1420-1425.

Lougheed, S., C. Gascon, D. Jones, J. Bogart, and P. Boag. 1999. Ridges and rivers: a test of competing hypotheses of Amazonian diversification using a dart-poison frog (*Epipedobates femoralis*). Proceedings of the Royal Society of London B 266: 1829-1835.

Lovejoy, N. R., S. C. Willis, and J. S. Albert. 2010. Molecular signatures of Neogene biogeographical events in the Amazon fish fauna Pages 405-417 in Amazonia, Landscape and Species Evolution: A Look into the Past (C. Hoorn and F. P. Wesselingh, eds.). Wiley-Blackwell, London, UK.

Lovette, I. J. 2004. Molecular phylogeny and plumage signal evolution in a trans-Andean and circum-Amazonian avian species complex. Molecular Phylogenetics and Evolution 32: 512-523.

Lundberg, J. G., M. H. Sabaj Pérez, W. M. Dahdul, and O. A. Aguilera. 2010. The Amazonian Neogene fish fauna. Pages 281-301 in Amazonia, Landscape and Species Evolution: A Look into the Past (C. Hoorn and F. P. Wesselingh, eds.). Wiley-Blackwell, London, UK.

Lynch, M., 2009. Estimation of allele frequencies from high-coverage genome-sequencing projects. Genetics 182: 295-301.

Maldonado-Coelho, M., J. Blake, L. Silveira, H. Batalha-Filho, and R. Ricklefs. 2013. Rivers, refuges and population divergence of fire-eye antbirds (*Pyriglena*) in the Amazon Basin. Journal of Evolutionary Biology 26: 1090-1107.

Mallet, J. 1993. Speciation, raciation, and color pattern evolution in *Heliconius* butterflies: evidence from hybrid zones. Pages 226-260 in Hybrid Zones and the Evolutionary Process (R. G. Harrison, ed.). Oxford University Press, Oxford, UK.

Mayr, E. 1963. Animal Species and Evolution. Belknap Press, Cambridge, USA.

Meirmans, P. G. 2012. The trouble with isolation by distance. Molecular Ecology 21: 2839-2846.

Mora, A., P. Baby, M. Roddaz, M. Parra, S. Brusset, W. Hermoza, and N. Espurt. 2010. Tectonic history of the Andes and sub-Andean zones: implications for the development of the Amazon drainage basin. Pages 38-60 in Amazonia, Landscape and Species Evolution: A Look into the Past (C. Hoorn and F. P. Wesselingh, eds.). Wiley-Blackwell, London, UK.

Moritz, C., J. L. Patton, C. J. Schneider, and T. B. Smith. 2000. Diversification of rainforest faunas: an integrated molecular approach. Annual Reviews in Ecology and Systematics 31: 533-563.

Nadeau, N. J., S. H. Martin, K. M. Kozak, C. Salazar, K. K. Dasmahapatra, J. W. Davey, S. W. Baxter, M. L. Blaxter, J. Mallet, and C. D. Jiggins. 2013. Genome-wide patterns of divergence and gene flow across a butterfly radiation. Molecular Ecology 22: 814-826.

Naka, L. N., C. L. Bechtoldt, L. M. P. Henriques, and R. T. Brumfield. 2012. The role of physical barriers in the location of avian suture zones in the Guiana Shield, northern Amazonia. American Naturalist 179: E115-E132.

Noonan, B.P. and P. Gaucher. 2005. Phylogeography and demography of Guianan harlequin toads (*Atelopus*): diversification within a refuge. Molecular Ecology 14: 3017-3031.

Nores, M. 1999. An alternative hypothesis for the origin of Amazonian bird diversity. Journal of Biogeography 26: 475-485.

Nosil, P., D. J. Funk, and D. Ortiz-Barrientos. 2009. Divergent selection and heterogeneous genomic divergence. Molecular Ecology 18: 375-402.

O'Donoghue, M., S. Boutin, C. J. Krebs, and E. J. Hofer. 1997. Numerical responses of coyotes and lynx to the snowshoe hare cycle. Oikos 80: 150-162.

Patton, J. L., M. N. F. da Silva, and J. R. Malcolm. 1994. Gene genealogy and differentiation among arboreal spiny rats (Rodentia: Echimyidae) of the Amazon basin: a test of the riverine barrier hypothesis. Evolution 48: 1314-1323.

Patton, J. L., M. N. F. da Silva, and J. R. Malcolm. 2000. Mammals of the Rio Juruá and the evolutionary and ecological diversification of Amazonia. Bulletin of the American Museum of Natural History 244: 1-306.

Pearson, D. L., 1977. A pantropical comparison of bird community structure on six lowland forest sites. Condor 79: 232-244.

Pennington, R. T. and C. W. Dick. 2010. Diversification of the Amazonian flora and its relation to key geological and environmental events: a molecular perspective. Pages 375-385 in Amazonia, Landscape and Species Evolution: A Look into the Past (C. Hoorn and F. P. Wesselingh, eds.). Wiley-Blackwell, London, UK.

Peterson, B. K., J. N. Weber, E. H. Kay, H. S. Fisher, and H. E. Hoekstra. 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and sequencing in model and non-model species. PLoS One 7: e37135.

Pinto, O. M. O. 1954. Resultados ornitológicos de duas viagens científicas ao Estado de Alagoas. Papéis Avulsos do Departamento de Zoologia, São Paulo 12: 1-98.

Pool, J. E., I. Hellmann, J. D. Jensen, R. Nielsen. 2010. Population genetic inference from genomic sequence variation. Genome Research 20: 291-300.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155: 945-959.

Quijada-Mascareñas, J. A., J. E. Ferguson, C. E. Pook, M. D. G. Salomão, R. S. Thorpe, and W. Wüster. 2007. Phylogeographic patterns of trans-Amazonian vicariants and Amazonian biogeography: the Neotropical rattlesnake (Crotalus durissus complex) as an example. Journal of Biogeography 34: 1296-1312.

Rambaut, A. and A. Drummond. 2007. Tracer v1.4. Available from http://www.tree.beast.bio.ed.ac.uk/software/tracer (last accessed 12 November 2015).

Rambaut, A. and A. Drummond. 2008. TreeAnnotator v1.4.8. Available from http://beast.bio.ed.ac.uk/TreeAnnotator (last accessed 12 November 2015).

Remsen, J. V. 2003. Family Furnariidae (Ovenbirds). Pages 162-357 in Handbook of the Birds of the World (J. del Hoyo, A. Elliott, J. Sargatal, and D. A. Christie, eds.). Lynx Edicions, Barcelona, Spain.

Ribas, C. C., A. Aleixo, A. C. Nogueira, C. Y. Miyaki, and J. Cracraft. 2012. A palaeobiogeographic model for biotic diversification within Amazonia over the past three million years. Proceedings of the Royal Society of London B 279: 681-689.

Ricardo Negri, F., J. Bocquentin-Villanueva, J. Ferigolo, and P. O. Antoine. 2010. A review of Tertiary mammal faunas and birds from western Amazonia. Pages 243-258 in Amazonia, Landscape and Species Evolution: A Look into the Past (C. Hoorn and F. P. Wesselingh, eds.). Wiley-Blackwell, London, UK.

Rull, V. 2013. Some problems in the study of the origin of Neotropical biodiversity using palaeoecological and molecular phylogenetic evidence. Systematics and Biodiversity 11: 415-423.

Sanmartín, I., P. Van Der Mark, and F. Ronquist. 2008. Inferring dispersal: a Bayesian approach to phylogeny-based island biogeography, with special reference to the Canary Islands. Journal of Biogeography 35: 428-449.

Santos, J. C., L. A. Coloma, K. Summers, J. P. Caldwell, R. Ree, and D. C. Cannatella. 2009. Amazonian amphibian diversity is primarily derived from late Miocene Andean lineages. PLoS Biology 7: e1000056.

Sclater, P. L. 1856. Catalogue of the birds collected by M. Auguste Sallé in southern Mexico, with descriptions of new species. Proceedings of the Zoological Society of London 24: 283-311.

Sclater, P. L. 1861. Descriptions of twelve new species of American birds of the families Dendrocolaptidae, Formicariidae, and Tyrannidae. Proceedings of the Zoological Society of London 29: 377-383.

Sick, H. 1967. Rios e Enchentes na Amazônia como Obstáculo para a Avifauna. Pages 495-520 in Atas do Simpósio sobre a Biota Amazônica (H. Lent, ed.). Conselho de Pesquisas, Rio de Janeiro, Brazil.

Sillett, T. S., R. T. Holmes, and T. W. Sherry. 2000. Impacts of a global climate cycle on population dynamics of a migratory songbird. Science 288: 2040-2042.

Smith, B. T., M. G. Harvey, B. C. Faircloth, T. C. Glenn, and R. T. Brumfield. 2014. Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. Systematic Biology 63: 83-95.

Smith, B. T., J. E. McCormack, A. M. Cuervo, M. J. Hickerson, A. Aleixo, C. D. Cadena, J. Pérez-Eman, C. W. Burney, X. Xie, M. G. Harvey, B. C. Faircloth, T. C. Glenn, E. P. Derryberry, J. Prejean, S. Fields, and R. T. Brumfield. 2014. The drivers of tropical speciation. Nature 515: 406-409.

Solomon, S. E., M. Bacci Jr., M. Martins Jr., G. G. Vinha, and U. G. Mueller. 2008. Paleodistributions and comparative molecular phylogeography of leafcutter ants (*Atta* spp.) provide new insight into the origins of Amazonian diversity. PLoS One 3: e2738.

Stapley, J., J. Reger, P. G. Feulner, C. Smadja, J. Galindo, R. Ekblom, C. Bennison, A. D. Ball, A. P. Beckerman, and J. Slate. 2010. Adaptation genomics: the next generation. Trends in Ecology and Evolution 25: 705-712.

Takahata, N., Y. Satta, and J. Klein. 1995. Divergence time and population size in the lineage leading to modern humans. Theoretical Population Biology 48: 198-221.

Teacher, A. G., C. André, P. R. Jonsson, and J. Merilä. 2013. Oceanographic connectivity and environmental correlates of genetic structuring in Atlantic herring in the Baltic sea. Evolutionary Applications 6: 549-567.

Todd, W. E. C. 1913. Preliminary diagnoses of apparently new birds from tropical America. Proceedings of the Biological Society of Washington 26: 169-174.

Tuomisto, H. and K. Ruokolainen. 1997. The role of ecological knowledge in explaining biogeography and biodiversity in Amazonia. Biodiversity and Conservation 6: 347–357.

Turner, T., M. McPhee, P. Campbell, and K. Winemiller. 2004. Phylogeography and intraspecific genetic variation of prochilodontid fishes endemic to rivers of northern South America. Journal of Fish Biology 64: 186-201.

Vieillot, L. P. 1816. Analyse d'une Nouvelle Ornithologie Élémentaire. Taylor and Francis, London UK.

Warren, W. C., D. F. Clayton, H. Ellegren, A. P. Arnold, L. W. Hillier, A. Künstner, S. Searle, S. White, A. J. Vilella, S. Fairley, A. Heger, L. Kong, C. P. Ponting, E. D. Jarvis, C. V. Mello, P. Minx, P. Lovell, T. A. F. Velho, M. Ferris, C. N. Balakrishnan, S. Sinha, C. Blatti, S. E. London, Y. Li, Y.-C. Lin, J. George, J. Sweedler, B. Southey, P. Gunaratne, M. Watson, K. Nam, N. Backström, L. Smeds, B. Nabholz, Y. Itoh, O. Whitney, A. P. Pfenning, J. Howard, M. Völker, B. M. Skinner, D. K. Griffin, L. Ye, W. M. McLaren, P. Flicek, V. Quesada, G. Velasco, C. Lopez-Otin, X. S. Puente, T. Olender, D. Lancet, A. F. A. Smit, R. Hubley, M. K. Konkel, J. A.Walker, M. A. Batzer, W. Gu, D. D. Pollock, L. Chen, Z. Cheng, E. E. Eichler, J. Stapley, J. Slate, R. Ekblom, T. Birkhead, T. Burke, D. Burt, C. Scharff, I. Adam, H. Richard, M. Sultan, A. Soldatov, H. Lehrach, S. V. Edwards, S.-P. Yang, X. Li, T. Graves, L. Fulton, J. Nelson, A. Chinwalla, S. Hou, E. R. Mardis, R. K. Wilson. 2010. The genome of a songbird. Nature 464: 757–762.

Weckstein, J. D. and R. Fleischer. 2005. Molecular phylogenetics of the *Ramphastos* toucans: Implications for the evolution of morphology, vocalizations, and coloration. Auk 122: 1191-1209.

Weir, J. T., 2006. Divergent timing and patterns of species accumulation in lowland and highland Neotropical birds. Evolution 60: 842-855.

Wesselingh, F. and J. Salo. 2006. A Miocene perspective on the evolution of the Amazonian biota. Scripta Geologica 133: 439-458.

White, T. A., S. E. Perkins, G. Heckel, and J. B. Searle. 2013. Adaptive evolution during an ongoing range expansion: the invasive bank vole (*Myodes glareolus*) in Ireland. Molecular Ecology 22: 2971-2985.

Wittmer, H. U., A. R. Sinclair, and B. N. McLellan. 2005. The role of predation in the decline and extirpation of woodland caribou. Oecologia 144: 257-267.

Zimmer, J. T. 1924. New birds from central Peru. Proceedings of the Field Museum of Natural History 12: 51-67.

Zimmer, J. T. 1935. Studies of Peruvian birds XVIII: diagnoses of new species and subspecies of Furnariidae from Peru and other parts of South America. American Museum. Novitates 819: 1-8.

CHAPTER 4: BIRDS OF UPLAND AND FLOODPLAIN FOREST IN THE AMAZON HAVE DIFFERENT EVOLUTIONARY HISTORIES

INTRODUCTION

Levels of geographic diversity vary widely across species (Taberlet et al. 1998, Soltis et al. 2006), in some cases because they have experienced different histories of landscape change (Lorenzen et al. 2012). In co-distributed species that have evolved under similar landscape histories, however, other factors may need to be invoked to explain differences in diversity (Lessios 2008). Although stochasticity may account for some of the variation, differences in the ecologies of species may have additional, deterministic effects on their evolutionary trajectories (Avise et al. 1987, Palumbi 1992).

Most evidence for deterministic ecological impacts on intraspecific histories comes from marine organisms. Larval dispersal mode in marine organisms may impact levels of population genetic structure, although the relationship is often complicated by the disparity between the ecological timescales on which dispersal occurs and those required to accrue divergence in many genetic markers (Palumbi 2003, Hellberg 2009). In terrestrial systems, Burney and Brumfield (2009) found that the forest stratum a bird inhabits predicts divergence across landscape barriers and Paz et al. (2015) found that body size, current landscape resistance, geographic range, biogeographic origin, and reproductive mode predicted divergence among areas in Panamanian frog species. Most prior estimates of geographic genetic divergence and population history, however, have been limited by the availability of independent genetic markers for parameter estimation.

New genomic approaches can be used to obtain genetic data from many independent parts of the genome and many samples (e.g., Davey et al. 2011, Faircloth et al. 2012). The number of independent loci in genomic datasets provides sufficient power to evaluate parameter-rich models of population history (Carstens et al. 2013). Further, increasing the number of loci in a dataset provides more precise estimates of parameter values that are less subject to biases resulting from disparities among gene histories (Edwards and Beerli 2000, Carling and Brumfield 2007). Concerted evolutionary responses to the ecological traits of species may be easier to detect when many, precise parameter estimates are available for examination. No study, however, has yet used comparisons of detailed estimates of geographic genetic diversity and historical parameters from genomic data to evaluate the impact of ecology on population history across species.

The avifauna of the Amazon Basin in northern South America is the most diverse in the world (Pearson 1977) and includes species with a variety of ecological traits (Parker et al. 1996) and variable levels of geographic genetic structure (Bates 2000, Smith et al. 2015). Many species are habitat specialists (Kratter 1997, Rosenberg 1990, Alonso et al. 2013) and closely related species often partition space by associating with different habitats. Two habitats in particular, floodplain forest (*várzea*) and upland forest (*terra firme*), are widespread and are inhabited by a suite of pairs of closely related species that segregate by habitat (Remsen and Parker 1983) and sometimes exhibit interspecific aggression (Robinson and Terborgh 1995). Floodplain forest receives nutrients from sediment-loaded whitewater rivers and also has an open, edge-like structure as a result of disturbance during floods (Prance 1979, Wittmann et al. 2004).

Upland forest, conversely, is often nutrient-deprived and is typified by a high proportion of tall trees, a dark interior, and open understory (Campbell et al. 1986, Gentry and Emmons 1987). Bird species of edge habitats, such as floodplain forest, exhibit more seasonal movements than species of forest interior (Levey and Stiles 1992) and floodplain forest species may be less limited by riverine barriers (Hayes and Sewlal 2004). These factors may result in higher gene flow and lower differentiation between populations that could explain the observation that subspecies richness is lower in floodplain than upland forest birds in Amazonia (Remsen and Parker 1983, Salisbury et al. 2012). Whether Amazonian birds exhibit habitat-associated differences in gene flow, geographic genetic diversity, and population history, however, is still unknown.

In this study, I examine forty species or species complexes (all of which are hereafter referred to as "species" for brevity) of widely co-distributed Amazonian birds that differ in habitat association. The forty species include twenty pairs in which one species is found in upland forest, and the other is a closely related species found in floodplain forest. I collect genomic sequence data from 2,416 ultraconserved elements and exons and use them to estimate genetic diversity, population structure, demographic history, and signals of selection in each species. I then test whether habitat preference predicts metrics of population diversity and history. I also test for an influence of the forest stratum a species inhabits and morphological metrics of dispersal ability on genetic parameters.

METHODS

Sample Design

We examined pairs of closely related species in an attempt to reduce bias resulting from differences in mutation rate, effective population size, or other processes that might vary in a concerted fashion among different clades in the avian tree of life. With the help of published data (Parker et al. 1996, del Hoyo et al. 2002-2011, Schulenberg et al. 2010) and expert knowledge (B. Whitney, Louisiana State University and L. Naka, Universidade Federal de Pernambuco), I selected genera that contained a pair of species or species complexes that generally segregate between floodplain and upland forest. I obtained lists of vouchered tissue samples collected during my fieldwork and available from natural history collections. From an initial list of 57 pairs that fit my criteria, I removed any pair containing a species for which fewer than 20 tissue samples were available in existing museum collections. I also removed pairs in which species boundaries or monophyly of populations with respect to other species were under debate (Remsen et al. 2015), or in which geographic breadth of sampling was insufficient to capture the entire Amazonian distribution of either member of the pair. The result was a list of 20 species pairs from 15 families.

For each species, I included all populations occurring within the Amazon, including allopatric populations that are considered separate species (species complexes). I removed any allopatric replacements, however, that were known to be distantly related to the remaining populations (e.g. *Xiphorhynchus pardalotus* appears to be the

geographic replacement of *X. elegans* in the Guianas, but is sister to another species, *X. ocellatus*). I georeferenced all records with locality information more precise than department/state and sufficient precision to determine on which side of any major biogeogeographic barriers (rivers or mountains) the sample originated. Locality records were plotted using ArcMap 10.0 (ESRI, Redlands, CA) with the WGS84 projection. I also digitized the Amazon terrestrial areas of endemism based on the limits in da Silva et al. (2005).

We designed geographic sampling in order to obtain the greatest similarity across taxa in the spatial dispersion of samples, thereby reducing bias in the results due to differences in sampling. I plotted random points across the Amazon using the genrandompts function in Geospatial Modelling Environment v. 0.7.1.0 (Beyer), with the minimum distance between points set to 2 map units (equivalent to two degrees in WGS84) and requiring 2 or more points within each area of endemism (da Silva 2005). For each species, I determined the closest sampling locality (linear distance) to each random point using the spatial join function in ArcMap. Some points are closest to the same sample, however, resulting in a list of fewer unique samples than points in each species. To reconcile this issue, I plotted more points initially (40) than required for the final sample (11), and arrived at the final sample by removing samples with low DNA concentrations (see below) or that were overly clustered geographically. Clustering was determined by projecting the samples on a grid using their coordinates and assessing clustering without referencing the underlying geography.

Many of the study species include populations outside of the Amazon Basin, either in the Atlantic Forest of southeastern South America or the humid forests of

Central America and the Chocó region of northwestern South America. For species with a population in either the Atlantic Forest or Central America and Chocó, I included a single sample from those regions in addition to the 20 Amazonian samples already selected to provide a larger geographical context for results.

Table 1. Study species.

	Genus	Upland Forest Species/Complex	Floodplain Forest Species/Complex
1	*Crypturellus*	*variegatus*	*undulatus*
2	*Piaya**	*melanogaster*	*cayana*
3	*Megascops*	*watsonii*	*choliba*
4	*Glaucidium**	*hardyi*	*brasilianum*
5	*Phaethornis*	*bourcieri/philippi*	*hispidus*
6	*Trogon**	*rufus*	*collaris*
7	*Monasa**	*morphoeus/atra*	*nigrifrons*
8	*Celeus**	*grammicus/undatus*	*flavus*
9	*Campephilus**	*rubricollis*	*melanoleucos*
10	*Myrmoborus*	*myotherinus*	*leucophrys*
11	*Myrmeciza*	*fortis*	*hyperythra*
12	*Hylophylax*	*naevius*	*punctulatus*
13	*Formicarius*	*colma*	*analis*
14	*Xiphorhynchus**	*elegans/spixii*	*obsoletus*
15	*Synallaxis*	*rutilans*	*gujanensis*
16	*Pipra*	*erythrocephala/rubrocapilla/chloromeros*	*filicauda/fasciicauda/aureola*
17	*Schiffornis*	*turdina*	*major*
18	*Pheugopedius/Cantorchilus*	*coraya/genibarbis*	*leucotis albipectus*
19	*Tachyphonus**	*cristatus*	*luctuosus*
20	*Saltator**	*grossus*	*coerulescens*

* = canopy species

Laboratory Methods

We extracted whole genomic DNA from tissues using DNeasy Blood and Tissue kits (Qiagen, Valencia, CA) and quantified extracts using a QuBit fluorometer

(ThermoFisher, Waltham, MA). I excluded samples with extracts containing less than 1 μg of DNA total. I removed additional samples based on spatial dispersion without reference to geography, as described above, to arrive at a final sample of 11 individuals per species plus extra-Amazonian outgroups.

Due to the comparative nature of my study, it was critical to obtain datasets in which genetic diversity and estimates of population history were unbiased across species. Results are generally not comparable across species if different loci are examined in different datasets and when settings used for orthology assessment among sequence reads leads to biased levels of variation across datasets (Harvey et al. 2015). Sequence capture of conserved genomic regions permits the interrogation of the same loci across divergent species (Faircloth et al. 2012, Bi et al. 2012, Hedtke et al. 2013), and orthology assessment in the assembly of sequence capture datasets is straightforward and has relatively little impact on allelic diversity (see Chapter 2).

We used sequence capture to target ultraconserved elements (UCEs) and exons from across the genome. I modified existing probe sets for UCEs (Faircloth et al. 2012) in order to obtain additional sequence from the more variable UCE flanking regions that might be useful for estimating shallow population histories. In UCE loci targeted with a single probe, I designed two probes extending further into the UCE flanks. The 120-mer probes were tiled such that they had 50% overlap (60 bp) in the middle of the locus and covered 180 bp total. Probe sequences were based on the chicken (*Gallus gallus*) genome release ICGSC Gallus_gallus-4.0 (Hillier et al. 2004). I also targeted conserved exons adjoining variable introns that have been used in previous avian phylogenetic studies (Kimball et al. 2009, Wang et al. 2012, Smith et al. 2012). Probes were designed off the

chicken genome sequence and were again tiled such that they covered the entire exon sequence at 2x coverage (50% overlap between adjoining probes). The final probe set included 4,715 probes targeting 2,321 UCEs and 96 exons.

We sent all samples to Rapid Genomics (Gainesville, FL) for sequence capture and sequencing following the general protocol described in Faircloth et al. (2012) and Smith et al. (2014). Samples were multiplexed at 160 samples per lane on a 100-bp paired-end Illumina HiSeq 2500 run. Rapid Genomics demultiplexed raw reads using custom scripts and strict barcode matching.

Bioinformatics

We cleaned reads with Illumiprocessor (Faircloth 2013). I developed a pipeline (https://github.com/mgharvey/seqcap_pop) to process and assemble datasets as follows. I used Velvet (Zerbino and Birney 2008) and the wrapper program Velvet Optimiser (Gladman 2009) to explore hash lengths between 67 and 71 and assemble reads across all individuals into contigs *de novo*. I mapped contigs to UCE probe sequences using Phyluce (Faircloth 2015). For each individual, I mapped reads to contigs that aligned to UCEs using bwa (Li and Durbin 2009). I explored thresholds that allowed anywhere from 1 to 7 mismatches between reads for mapping and found that the loss of alleles plateaued in many species at 4 mismatches per read, so I selected that setting for the final assembly. I converted sam files to bam format using samtools (Li et al. 2009) and cleaned bam files by soft-clipping reads outside the reference contigs with PICARD (http://broadinstitute.github.io/picard/). I added read groups for each individual using

PICARD and merged the bam files across individuals with samtools. I realigned reads to minimize mismatched bases using the RealignerTargetCreator and realigned indels using IndelRealigner in the GATK (McKenna et al. 2010). I called single nucleotide polymorphisms (SNPs) and indels using the GATK UnifiedGenotyper, annotated SNPs with VariantAnnotator, and masked indels using VariantFiltration. I removed SNPs with a quality score below Q30 and conducted read-backed phasing using the GATK. I extracted SNPs in vcf format and used add_phased_snps_to_seqs_filter.py from the seqcap_pop pipeline to insert SNPs into reference sequences and produce alignments for each locus across individuals. SNPs on the same locus for which phasing failed were inserted using the appropriate IUPAC ambiguity codes. I collated sequences and produced final alignments using MAFFT (Katoh et al. 2005).

We also assembled partial mitochondrial genomes for each sample from off-target reads using a similar pipeline. I obtained existing complete or nearly complete mitochondrial genome sequences from the most closely related taxon to each study species/complex for which they were available (Appendix C). I mapped reads to the mitochondrial genomes, sorted the bam file, recalculated MD tags, and indexed the bam file using Samtools. I then called variant sites and output vcf files containing variant and invariant bases using Freebayes (Garrison and Marth 2012) and used these to assemble sequences using freebayes_vcf2fa_mt.py (https://github.com/mgharvey/misc_python/bin/freebayes_vcf2fa.py). Only sites with a read depth of 5 or greater were included in sequences. I conducted final alignment with MAFFT.

We searched for potential sample identification errors or signs of contamination by building exploratory trees of concatenated SNPs from the UCE/exon data using

MrBayes v.3.2.2 (Ronquist et al. 2013) and scrutinizing any long branches and by mapping mitochondrial sequences to existing sequence data in Genbank (Benson et al. 2014) using Blastn (Altschul et al. 1997). I counted the reads in BWA assemblies using Samtools. I calculated basic population genetic summary statistics including nucleotide diversity (π)(Tajima 1983) and Watterson's θ (Watterson 1975) across all ingroup samples in each species using DendroPy v.3.10.0 (Sukumaran and Holder 2010).

Population Structure Inference

Differences in ecological traits between floodplain and upland forest may drive differences in levels of population structure between the habitats by impacting gene flow, rates of neutral divergence, and selection. I inferred population structure using the ingroup samples from each species/complex. Diverse methods are available to infer population structure (reviewed in Leavitt et al. 2015), and they can provide different results (Latch et al. 2006, Chen et al. 2007). I examined three methods to infer population structure and individual population assignments: STRUCTURE (Pritchard et al. 2000), Bayesian Analysis of Population Structure (BAPS; Corander et al. 2003), and Discriminant Analysis of Principal Components (DAPC; Jombart et al. 2010). I also used the first two methods to infer admixture in each individual.

STRUCTURE is a model-based clustering method that simultaneously infers population structure and assesses the probability of individual assignment to a cluster or combination of clusters (admixture) at a given number of clusters (K). Individuals are assigned to clusters so as to minimize Hardy-Weinberg and linkage disequilibrium. I ran

STRUCTURE using the linkage model, and provided phase information for each site in each individual as well as distances in base-pairs between linked sites. Sites mapping to different loci were treated as unlinked. I conducted analyses at K values ranging from 1 to 6, with 10 replicate runs at each value. Each run included a 50,000-iteration burn-in followed by 200,000 sampling iterations, and I assessed convergence by examining alpha, F, D_{ij}, and the likelihood within and across runs at each K value. I estimated the best value of k using the method of Evanno et al. (2005) implemented in StructureHarvester (Earl 2012). In some cases, the results at the best K value included clusters to which no individuals were assigned. In these situations, I also examined the largest K value in which at least one individual was assigned to each cluster. I combined results across replicates run with the best K value using CLUMPP (Jakobsson and Rosenberg 2007).

BAPS is a model-based clustering method similar to STRUCTURE in that it clusters individuals so as to minimize Hardy-Weinberg and linkage disequilibrium. Although the inference of number of clusters (K) in STRUCTURE is determined by *ad hoc* interpretation of the posterior probability of K from separate MCMC chains, BAPS assesses K using a greedy stochastic optimization algorithm. After population assignments assessed, admixture can be assessed using a subsequent, simulation-based analysis. Because BAPS requires complete phasing information for linked sites, and phasing had failed for some individuals at most linked sites in my datasets, I used the unlinked model and examined only a single randomly selected SNP from each locus for this analysis. I conducted mixture clustering with the maximum number of populations (k) set at 10. I estimated admixture in each individual based on mixture clustering using

50 simulation iterations, 50 reference individuals, and 10 iterations to estimate admixture coefficients in the reference individuals.

DAPC is a fast, multivariate method for inferring the number of genetic clusters and cluster assignments in large datasets. I inferred the number of clusters and cluster membership in DAPC using the maximum number of PCs available for each species, and selected the best value for cluster number by choosing the value at which Bayesian Information Criterion reached a low point (Jombart et al. 2010). Unlike STRUCTURE and BAPS, DAPC does not allow for admixture estimation.

We compared levels of inferred population structure across species and used population assignments from all three methods to conduct subsequent population-based analyses. I estimated F_{st} among populations and θ for each population using DendroPy and conducted an AMOVA across populations on each dataset. I also obtained a simple summary of structure across the distribution of each species by calculating the ratio of mean genetic distance between alleles in different individuals versus mean genetic distance between the two alleles within each individual using a custom script. This metric measures the distance between all individuals in a species, each of which comes from a different geographic region, while controlling for within-population diversity, analogous to popular estimators of population genetic F-statistics. Distances were corrected for multiple substitutions using the method of Jukes and Cantor (1969).

Demographic Modeling

Floodplain and terra firme species may differ in population size or other demographic parameters owing to differences in the structure, distribution, and landscape history of their preferred habitat. I estimated demographic parameters using a coalescent modeling approach in G-PhoCS v.1.2.3 (Gronau et al. 2011). I ran analyses using all population assignments from STRUCTURE, BAPS, and DAPC to assign population membership and specified the population topologies in situations where more than two populations were present based on Bayesian phylogenetic estimates using concatenated SNPs. For each species, I examined both a model with no migration between populations subsequent to divergence as well as a model allowing for asymmetric migration between all terminal populations. The parameters estimated were mutation-scaled effective population size ($\theta = 4N\mu$, where N is number of individuals and μ is mutation rate) for each ancestral and contemporary population, divergence time ($\tau = time*\mu$) between populations, and migration rate ($M = m\theta/4$, where m is the instantaneous rate of migration) between contemporary populations. I used gamma priors of (1, 5000) for θ and τ and (1,3) for migration and conducted runs of at least 500,000 iterations (sampling every 100). I also explored the impact of θ and τ priors of (1, 50). Convergence was assessed by examining parameter traces and ESS values in Tracer v.1.5 (Rambaut and Drummond 2007). G-PhoCS implements a multi-population model and cannot be run in the study species with a single population. For comparative analyses across species, I used the ingroup-wide θ values as calculated using DendroPy and divergence time (τ) values of zero for single-population species.

Selection

Natural and sexual selection may drive divergence between populations, and differences in the strength of selection could underlie any differences in population genetic divergence between floodplain and upland forest species. Sexually selected traits often map to sex chromosomes (Reinhold 2008), and fast rates of evolution on sex chromosomes compared to autosomes are sometimes considered evidence for sexual selection (Counterman et al. 2004). I conducted a preliminary examination of relative rates of evolution on the Z chromosome and autosomes in my study species. I mapped contigs from each locus in each species to the Zebra Finch genome (Warren et al. 2010), the closest relative to most of my study species for which chromosome assemblies are available. In each study species I calculated the proportion of loci with fixed SNPs among BAPS populations on the Z chromosome and on the autosomes. Fixed SNPs were any SNPs with alternate alleles fixed between at least two BAPS populations. Sites with more than 50% missing data in the populations under consideration were not included. I also examined the relative depths of gene trees in expected substitutions per site between the Z chromosome and autosomes. Gene trees were estimated in RAxML v.8 (Stamatakis 2014) for each species.

Comparative Analyses

We assessed whether habitat predicted various metrics of population genomic diversity and population history using generalized linear models with mixed effects (GLMMs). The generalized linear modeling approach allowed us to examine response variables with diverse error distribution models in the same statistical framework. Gaussian error models were used for continuous and large count data, Poisson models for data composed of low count values (<100), and Gamma models with a logarithmic link function for continuous data with positive skew. The use of mixed effects allowed us to simultaneously consider habitat as a predictor variable and, to account for covariance due to shared history between species pairs, genus as a random variable. I examined the relationship between habitat and each genetic response variable in one-way tests using functions for GLMMs in the stats R package (R Core Team 2015).

Covariance due to shared history can also be modeled using a phylogenetic control. I estimated a phylogeny for the study species by aligning UCE and exon sequences from one sample of each species in MAFFT. Because sequences were assembled by mapping to different contigs in each species, the sequences were generally not entirely overlapping across species, and these ragged ends frequently included messy and potentially spuriously aligned blocks of sites. I removed these by filtering for only sites without missing data in the alignment. I concatenated filtered alignments that contained all 40 individuals and conducted Bayesian analyses on the complete matrix in MrBayes. I square-root transformed right-skewed variables to achieve normality and conducted phylogenetic generalized least squares (PGLS) analysis comparing habitat

association to metrics of geographic genetic divergence using the caper package (Orme et al. 2013) in R.

Although my primary focus was on the associations between habitat and genetic diversity, I also examined two additional traits thought to predict population divergence in Neotropical birds. First, whether a bird inhabits the forest canopy or understory (microhabitat) has been shown to predict levels of divergence across landscape barriers (Burney and Brumfield 2009, Smith et al. 2015), so I tested whether canopy and understory species (based on Parker et al. 1996) differed in metrics of population genomic diversity. Second, differences in habitat or microhabitat associations may influence population genetic divergence via differences in dispersal ability. I examined whether Kipp's Index, a morphological index of dispersal ability that can be measured from museum specimens (Kipp 1959), predicted levels of population genomic diversity across species. I again treated genus as a random variable and treated forest stratum and Kipp's index as fixed variables in one-way GLMM analyses. I also examined GLMMs with all three possible combinations of multiple predictor variables (habitat, forest stratum, and Kipp's index) to account for second-order interactions. Similarly, I ran additional PGLS analyses using forest stratum, Kipp's index, and all possible combinations of multiple predictor variables.

RESULTS

We obtained a mean value of 2,087,266 (s.d. = 656,446) raw reads per sample. A mean value of 28.1% (s.d. = 6.57%) of sequence reads were successfully mapped to

target loci after cleaning. A mean value of 0.44% (s.d. = 0.60%) of all reads mapped successfully to the mitochondrion. Across species, I obtained data from a mean of 2,142 UCEs (s.d. = 65.5) and 69 exons (s.d. = 4.8). I recovered data in at least one species from 2,416 of 2,417 targeted loci. Mean alignment length was 554 bp (s.d. = 56.3), and there were a mean of 7,196 (s.d. = 1,379) sites that were variable within the ingroup of each species. Additional summary statistics are provided in Appendices 1 and 2.

Eight samples appeared to be misidentified or heavily contaminated and were removed from further analyses (Appendix C). These samples had very long branches in Bayesian phylogenetic trees of concatenated SNPs and mitochondrial sequences from these samples mapped to distant relatives. Three samples contained large numbers of rare alleles likely to be a result of lower levels of contamination or sequencing errors and were also removed (Appendix C). Three samples failed, with greater than 85% missing data at variable sites, and were also removed (Appendix C).

The number of populations and population assignments inferred from STRUCTURE, BAPS, and DAPC were broadly concordant (Figs. 1, Appendix C). The best k-value from STRUCTURE analyses based on the Evanno method, both before and after reducing k to remove clusters without assigned individuals, ranged between one and four across study species (median = 3). The number of populations estimated in BAPS varied from one to three (median = 2), and in the number of clusters from DAPC varied between one and four (median = 2). Mean genetic distance between individuals varied from 1.13 to 3.13 times greater than between chromosomes within individuals (mean = 1.55). Many individuals contained non-zero probabilities of assignment to multiple

clusters in the STRUCTURE results, potentially indicative of admixture, but no admixture was recovered in the admixture analysis from BAPS.

Because standard estimates of mutation rate are not available for the sequenced UCE and exon loci, I examined raw estimates of mutation-scaled effective population size (θ; in units of size*μ), divergence time (τ; in units of time*μ), and migration rate (M; in units of individuals/μ) from G-PhoCS for the 23 species with multiple populations (Fig. 1, Appendix C). The mean per-site mutation-scaled effective population size in contemporary populations was 1.53×10^{-3} (s.d. = 5.73×10^{-4}). In contemporary populations, θ averaged 2.68 times larger than the θ inferred for the ancestral population at the root (s.d. = 1.29). Divergence time varied from 9.72×10^{-5} to 1.13×10^{-3} (mean = 4.45^{-4}). Average migration rate between populations within a species varied from 0.337 to 4.69 (mean = 0.950) and was positively correlated with τ estimates within a species (r^2 = 0.477, $p < 0.001$).

Across study species, contigs from 2,415 of 2,416 recovered loci successfully mapped to the Zebra Finch genome assembly. Contigs from all species mapped to the Z chromosome for 171 loci, to one of the autosomes for 2,169 loci, and to unplaced scaffolds in 44 loci. For 31 loci, contigs from different species mapped to different chromosomes or scaffolds resulting in ambiguous positions.

The average depths of gene trees across ingroup individuals were similar on the Z chromosome and the autosomes (Z chromosome gene trees averaged 1.07 times deeper, s.d. = 0.29). The Z chromosome, however, had an average of 2.53 times (s.d. = 1.78) more loci with SNPs that were fixed among BAPS populations than the autosomes.

Habitat association predicted three metrics of population genomic diversity in single-comparison GLMM analyses (Fig. 2, top). The number of variable sites among ingroup samples was greater in upland forest species ($t = 2.25$, $p = 0.031$), gene trees

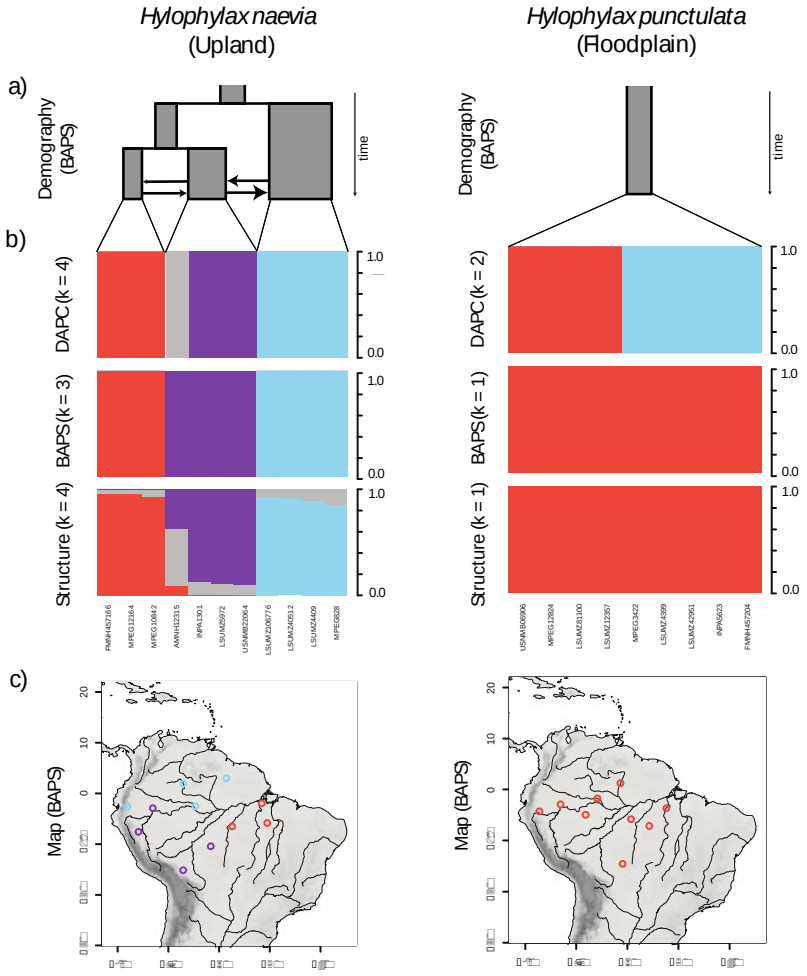


Figure. 1. Representative graphic of demographic models, population structure across three methods, and maps showing BAPS population distributions for one pair of study species.

averaged deeper in upland forest species (t = 2.39, p = 0.022), and the mean genetic distance between versus within individuals was greater in upland species (t = 2.64, p = 0.012). Forest stratum was associated with three variables (Fig. 2, bottom), the mean genetic distance between versus within individuals (t = 3.15, p = 0.003), the deepest divergence in the demographic model (t = 2.31, p = 0.032), and the relative gene tree depths between the Z chromosome and the autosomes (t = 2.98, p = 0.005). Kipp's index was associated with one variable, population size increase (the increase in θ between the root population and contemporary populations) in demographic analyses (t = -2.25, p =

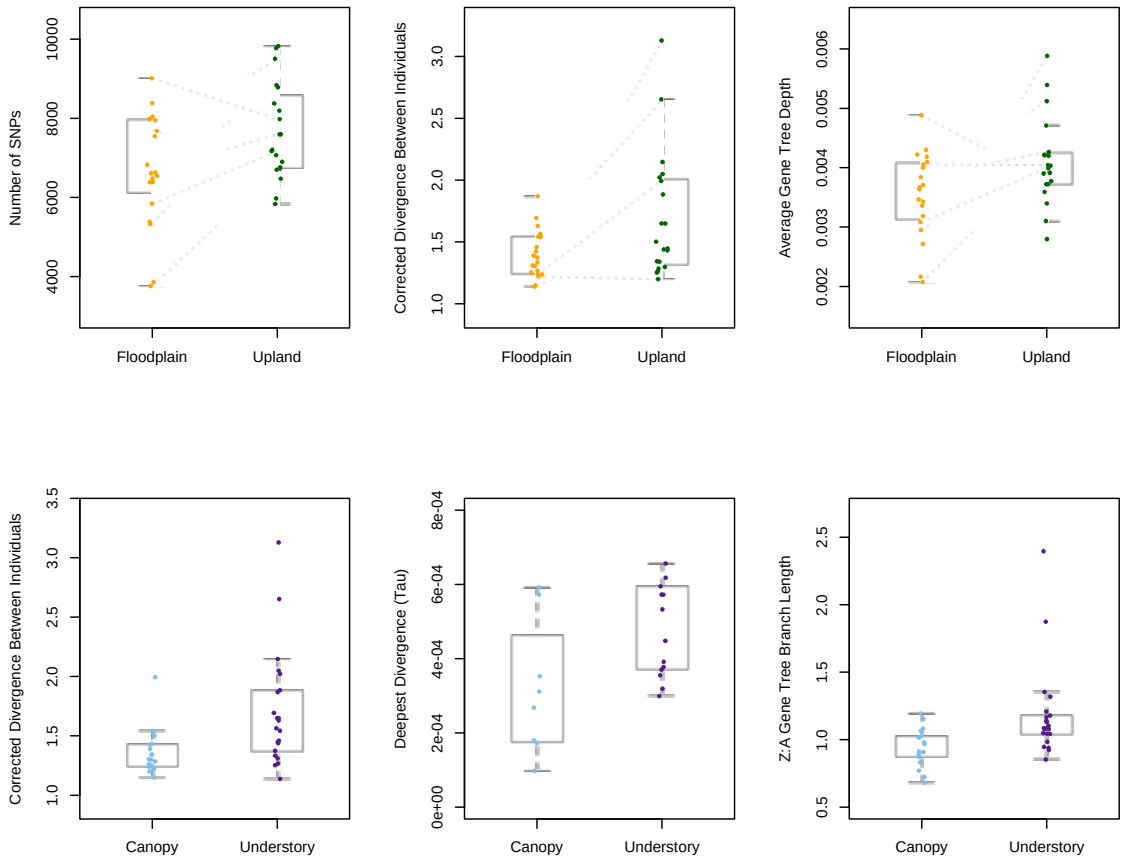


Figure 2. Differences in some population genetic parameters between floodplain and upland forest bird species and canopy and understory bird species. Dotted lines in the floodplain-upland plots connect paired members of the same genus.

0.031). These relationships changed little when considering second-order effects in multiple-predictor tests (Appendix C).

PGLS results indicated that, as with the GLMM, habitat predicted the relative genetic distance between versus within individuals ($t = 2.15$, $p = 0.044$). The number of populations from STRUCTURE, after reducing k to remove clusters with no assigned individuals, also was related to habitat ($t = 2.25$, $p = 0.036$), as was the number of DAPC clusters ($t = 2.74$, $p = 0.012$). Neither forest stratum nor Kipp's index was correlated with any response variable based on PGLS in one-way comparisons, although additional correlations did emerge in analyses with multiple predictor variables (Appendix C).

DISCUSSION

We found that the habitat associations of Amazonian birds predict diverse metrics of genetic diversity and population history, including divergence between individuals, average gene tree depth and levels of population structure (Fig. 2, top). Diversity and divergence were higher in upland forest than in floodplain forest in all significant comparisons. Several historical reasons may account for the observed disparity in diversity and population history across habitats. Dispersal may be greater in bird species of floodplain forest, leading to increased gene flow between populations that retards divergence (Salisbury et al. 2012). Seasonal flooding may force floodplain forest birds into upland forest, promoting colonization of new areas (Rosenberg 1990). Rivers, important barriers to dispersal in Amazonia, could be less effective dispersal barriers to floodplain species than to upland species (Capparella 1987, Patton and da Silva 1998).

Dispersal may be facilitated because floodplain species on opposite banks of a river are in greater proximity than upland forest species, because upland forest may not occur within several km of the main channel (Melack and Hess 2011). Moreover, river capture events may regularly shift patches of floodplain forest (Salo et al. 1986, Dumont 1991) and associated organisms (Tuomisto and Ruokolainen 1997, Patton et al. 2000) from one side of a river to another, but river capture events involving upland forest are likely very rare. I did not, however, recover higher migration rates in floodplain forest species than in upland forest species, nor did my morphological measure of dispersal ability differ between floodplain and upland species. my data do not suggest that higher dispersal between areas is the source of lower diversity and population structure in floodplain bird species.

Population demography may also result in different levels of diversity among habitats. Floodplains are relatively restricted in the Amazon Basin, where they cover about 14% of the lowland area (Melack and Hess 2011). The small area in floodplains may result in smaller effective population sizes in floodplain species, leading to lower genetic diversity and fewer opportunities for population divergence. Since θ scales with effective population size, correlations between θ measures and habitat might indicate demographic differences between floodplain and upland forest species, but I recovered no significant relationships between habitat and θ measures.

Low genetic diversity in floodplain forest bird species may be a result of recent expansions from historical bottlenecks (Aleixo 2006, Matocq et al. 2000). I found negative average Tajima's D values across loci, consistent with population expansion, but this pattern could also be due to purifying selection, which is thought to act on conserved

loci like those that I examined (Katzman et al. 2007). Demographic modeling also suggested that present-day populations were larger than ancestral root populations, but this may also reflect biases in the frequency of older polymorphisms due to purifying selection. Neither Tajima's D or increase in θ between the root and the present-day populations differed between floodplain and upland forest species. Recent colonization from other habitats or regions could leave similar genetic signatures to recent population expansion, although at least some floodplain species represent old lineages (Aleixo 2002). Many of my study taxa were closely related to samples from outside of the Amazon Basin, but I lacked sufficient sampling for a detailed reconstruction of their biogeographic history. More research will be required, possibly including more species and improved methods of modeling population history, to tease apart the processes underling differences in diversity and population history between floodplain and upland forest bird species.

The forest stratum at which a species is found was also related to some genetic metrics of divergence, consistent with prior results from comparative studies in Neotropical birds (Burney and Brumfield 2009, Smith et al. 2015). These correlations were recovered despite the fact that my sampling, designed to maximize independent comparisons across habitat types, was not optimal for detecting an impact of forest stratum (with roughly half the number of independent samples). The forest canopy is in many ways analogous to edge habitats like floodplain forest, and both are thought to harbor higher concentrations of birds that undergo seasonal movements than tall forest (Levey and Stiles 1992). Both canopy and floodplain bird species have lower subspecies richness than understory and upland forest species, respectively (Salisbury et al. 2012).

The difference between canopy and understory species in the branch length ratio between the Z chromosome and autosomes seems surprising, but appears to be driven by two species: both tinamous have relatively long branches on the Z chromosome and inhabit the understory. The relationship between Kipp's index and increases in θ over time, with species with morphologies associated with poorer dispersal abilities exhibiting higher rates of increase, may be worth further investigation.

The metrics of population genetic diversity and divergence selected here are certainly not the only, or even the best, measures available to examine genetic impacts of habitat or other ecological variables. Measures of population structure based on STRUCTURE, BAPS, and DAPC only showed significant relationships with predictor variables in a few cases (STRUCTURE and DAPC results in PGLS analysis versus habitat). This may be partly due to the small variance in structure estimates based on these programs (K values ranged only from 1 to 4 populations). The continuous measure of between versus within individual divergence appeared to show stronger relationships with the predictor variables. Finer geographic sampling or methods that are able to uncover finer-scale population structure might provide sufficient variation in structure estimates to recover patterns in comparative analyses. The development of other methods for estimating detailed population genetic diversity or population history based on genomic datasets may also reveal patterns.

We have demonstrated that ecological traits, in particular habitat associations, predict diverse population genetic differences across species. Interestingly, the upland forest avifauna is more diverse (1,058 species) than the floodplain forest avifauna (154 species)(Parker et al. 1996). Given the association between population divergence and

speciation rate over long evolutionary timescales (see Chapter 5), different rates of population divergence between upland and floodplain forest may have played a role in producing their disparate diversities, a process known as species selection (Stanley 1975). I have also demonstrated that genomic datasets can be used to estimate diverse parameters for testing hypotheses about the factors associated with genomic diversity. Studies examining additional taxa and new methods for estimating more detailed population histories are sure to provide more insight into the impacts of ecology on population genomics and evolution in the near future.

REFERENCES

Aleixo, A. 2006. Historical diversification of floodplain forest specialist species in the Amazon: a case study with two species of the avian genus *Xiphorhynchus* (Aves: Dendrocolaptidae). Biological Journal of the Linnaean Society 89: 383-395.

Alonso, J. Á., M. R. Metz, and P. V. Fine. 2013. Habitat specialization by birds in western Amazonian white-sand forests. Biotropica 45: 365-372.

Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and J. D. Lipman. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Research 25: 3389-3402.

Avise, J. C., J. Arnold, R. M. Ball, E. Bermingham, T. Lamb, J. E. Neigel, C. A. Reeb, and N. C. Saunders. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annual Review of Ecology and Systematics 18: 489-522.

Bates, J. M. 2000. Allozymic genetic structure and natural habitat fragmentation: data for five species of Amazonian forest birds. Condor 102: 770-783.

Benson, D. A., K. Clark, I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, and E. W. Sayers. 2014. GenBank. Nucleic Acids Research 42: D32-37.

Bi, K., D. Vanderpool, S. Singhal, T. Linderoth, C. Moritz, and J. Good. 2012. Transcriptome-based exon capture enables highly cost-effective comparative genomic data collection at moderate evolutionary scales. BMC Genomics 13: 403.

Burney, C. W. and R. T. Brumfield. 2009. Ecology predicts levels of genetic differentiation in Neotropical birds. American Naturalist, 174: 358-368.

Campbell, D. G., D. C. Daly, G. T. Prance, and U. N. Maciel. 1986. Quantitative ecological inventory of terra firme and várzea tropical forest on the Rio Xingu, Brazilian Amazon. Brittonia 38: 369-393.

Capparella, A. P. 1987. Effects of riverine barriers on genetic differentiation of Amazonian forest undergrowth birds. PhD Dissertation. Louisiana State University, Baton Rouge, LA.

Carling, M. D., and R. T. Brumfield. 2007. Gene sampling strategies for multi-locus population estimates of genetic diversity (θ). PLoS One 2: e160.

Carstens, B. C., R. S. Brennan, V. Chua, C. V. Duffie, M. G. Harvey, R. A. Koch, C. D. McMahan, B. J. Nelson, C. E. Newman, J. D. Satler, G. Seeholzer, K. Posbic, D. C. Tank, and J. Sullivan. 2013. Model selection as a tool for phylogeographic inference: an example from the willow Salix melanopsis. Molecular Ecology 22: 4014-4028.

Chen, C., E. Durand, F. Forbes, and O. François. 2007. Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. Molecular Ecology Notes 7: 747-756.

Corander, J., P. Waldmann, and M. J. Sillanpää. 2003. Bayesian analysis of genetic differentiation between populations. Genetics 163: 367-374.

Counterman, B. A., D. Ortiz-Barrientos, and M. A. F. Noor. 2004. Using comparative genomic data to test for fast-X evolution. Evolution 58: 656-660.

Da Silva, J. M. C., A. B. Rylands, and G. A. B. Da Fonseca. 2005. The fate of the Amazonian areas of endemism. Conservation Biology 19: 689-694.

Davey, J. W., P. A. Hohenlohe, P. D. Etter, J. Q. Boone, J. M. Catchen, and M. L. Blaxter. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nature Reviews Genetics 12: 499-510.

del Hoyo, J., A. Elliott, J. Sargatal, and D. Christie (eds.). 1992-2011. Handbook of the Birds of the World. Lynx Edicions, Barcelona, Spain.

Dumont, J.-F. 1991. Cambio del curso de los ríos Marañon y Ucayali en la depresión Ucamara durante el Cuaternario Reciente. Pages 483–487 in Volumen de Resúmenes Extendidos, VII Congreso Peruano de Geologia, Lima, Peru.

Earl, D. A. 2012. Structure Harvester: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources 4: 359-361.

Edwards, S. and P. Beerli. 2000. Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. Evolution 54: 1839-1854.

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology 14: 2611-2620.

Faircloth, BC. 2013. illumiprocessor: a trimmomatic wrapper for parallel adapter and quality trimming. Available from http://dx.doi.org/10.6079/J9ILL (last accessed 20 September 2015).

Faircloth, B. C. 2015. PHYLUCE is a software package for the analysis of conserved genomic loci. Bioinformatics doi:10.1093/bioinformatics/btv646.

Faircloth, B. C., J. E. McCormack, N. G. Crawford, M. G. Harvey, R. T. Brumfield, and T. C. Glenn. 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. Systematic Biology 61: 717-726.

Garrison, E. and G. Marth. 2012. Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:1207.3907.

Gentry, A. H. and L. H. Emmons. 1987. Geographical variation in fertility, phenology, and composition of the understory of Neotropical forests. Biotropica 19: 216-227.

Gladman, S. 2009. Velvet Optimiser. Available from http://bioinformatics.net.au/software.velvetoptimiser.shtml (last accessed December 20, 2013).

Gronau, I., M. J. Hubisz, B. Gulko, C. G. Danko, and A. Siepel. 2011. Bayesian inference of ancient human demography from individual genome sequences. Nature Genetics 43: 1031-1034.

Hayes, F. E. and J. A. N. Sewlal. 2004. The Amazon River as a dispersal barrier to passerine birds: effects of river width, habitat and taxonomy. Journal of Biogeography 31: 1809-1818.

Hedtke, S. M., M. J. Morgan, D. C. Cannatella, and D. M. Hillis. 2013. Targeted enrichment: maximizing orthologous gene comparisons across deep evolutionary time. PLoS One 8: e67908.

Hellberg, M. E., R. S. Burton, J. E. Neigel, and S. R. Palumbi. 2002. Genetic assessment of connectivity among marine populations. Bulletin of Marine Science 70: 273-290.

Hillier W. H., W. Miller, E. Birney, W. Warren, R. C. Hardison, C. P. Ponting, P. Bork, D. W. Burt, M. A. M. Groenen, M. E. Delaney, J. B. Dodgson, A. T. Chinwalla, P. F. Cliften, S. W. Clifton, K. D. Delehaunty, C. Fronick, R. S. Fulton, T. A. Graves, C. Kremitzi, D. Layman, V. Magrini, J. D. McPherson, T. L. Miner, P. Minx, W. E. Nash, M. N. Nhan, J. O. Nelson, L. G. Oddy, C. S. Pohl, J. Randall-Maher, S. M. Smith, J. W. Wallis, S.-P. Yang, M. N. Romanov, C. M. Rondelli, B. Paton, J. Smith, D. Morrice, L. Daniels, H. G. Tempest, L. Robertson, J. S. Masabanda, D. K. Griffin, A. Vignal, V. Fillon, L. Jacobbson, S. Kerje, L. Andersson, R. P. M. Crooijmans, J. Aerts, J. J. van der Poel, H. Ellegren, R. B. Caldwell, S. J. Hubbard, D. V. Grafham, A. M. Kierzek, S. R. McLaren, I. M. Overton, H. Arakawa, K. J. Beattie, Y. Bezzubov, P. E. Boardman, J. K. Bonfield, M. D. R. Croning, R. M. Davies, M. D. Francis, S. J. Humphray, C. E. Scott, R. G. Taylor, C. Tickle, W. R. A. Brown, J. Rogers, J.-M. Buerstedde, S. A. Wilson, L. Stubbs, I. Ovcharenko, L. Gordon, S. Lucas, M. M. Miller, H. Inoko, T. Shiina, J. Kaufman, J. Salomonsen, K. Skjoedt, G. K.-A. Wong, J. Wang, B. Lio, J. Wang, J. Yu, H. Yang, M. Nefedov, M. Koriabine, P. J. deJong, L. Goodstadt, C. Webber, N. J. Dickens, I. Letunic, M. Suyama, D. Torrents, C. von Mering, E. M. Zdobnov, K. Makova, A. Nekrutenko, L. Elnitski, P. Eswara, D. C. King, S. Yang, S. Tyekucheva, A. Radakrishnan, R. S. Harris, F. Chiaromonte, J. Taylor, J. He, M. Rijnkels, S. Griffiths-Jones, A. Ureta-Vidal, M. M. Hoffman, J. Severin, S. M. J. Searle, A. S. Law, D. Speed, D. Waddington, Z. Cheng, E. Tuzun, E. Eichler, Z. Bao, P. Flicek, D. D. Shteynberg, M. R. Brent, J. M. Bye, E. J. Huckle, S. Chatterji, C. Dewey, L. Pachter, A. Kouranov, Z. Mourelatos, A. G. Hatzigeorgiou, A. H. Paterson, R. Ivarie, M. Brandstrom, E. Axelsson, N. Backstrom, S. Berlin, M. T. Webster, O. Pourquie, A. Reymond, C. Ucla, S. E. Antonarakis, M. Long, J. J. Emerson, E. Betrán, I. Dupanloup, H. Kaessmann, A. S. Hinrichs, G. Bejerano, T. S. Furey, R. A. Harte, B. Raney, A. Siepel, W. J. Kent, D. Haussler, E. Eyras, R. Castelo, J. F. Abril, S. Castellano, F. Camara, G. Parra, R. Guigo, G. Bourque, G. Tesler, P. A. Pevzner, A. Smit, L. A. Fulton, E. R. Mardis, R. K. Wilson. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432: 695-716.

Jakobsson, M. and N. A. Rosenberg. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23: 1801-1806.

Jombart, T., S. Devillard, and F. Balloux. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genetics, 11: 94.

Jukes, T. H. and C. R. Cantor. 1969. Evolution of protein molecules. Mammalian Protein Metabolism 3: 21-132.

Katoh, K., K. I. Kuma, H. Toh, and T. Miyata. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Research 33: 511-518.

Katzman, S., A. D. Kern, G. Bejerano, G. Fewell, L. Fulton, R. K. Wilson, S. R. Salama, and D. Haussler. 2007. Human genome ultraconserved elements are ultraselected. Science 317: 915-915.

Kipp, F. A. 1959. Der Handflügel-Index als flugbiologisches Maβ. Die Vogelwarte 20: 77–86.

Kimball, R. T., E. L. Braun, F. K. Barker, R C. Bowie, M. J. Braun, J. L. Chojnowski, S. J. Hackett, K.-L. Han, J. Harshman, V. Heimer-Torres, W. Holznagel, C. J. Huddleston, B. D. Marks, K. J. Miglia, W. S. Moore, S. Reddy, F. H. Sheldon, J. V. Smith, C. C. Witt, and T. Yuri. 2009. A well-tested set of primers to amplify regions spread across the avian genome. Molecular Phylogenetics and Evolution 50: 654-660.

Kratter, A. W. 1997. Bamboo Specialization by Amazonian Birds. Biotropica 29: 100-110.

Latch, E. K., G. Dharmarajan, J. C. Glaubitz, and O. E. Rhodes, Jr. 2006. Relative performance of Bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. Conservation Genetics 7: 295-302.

Lessios, H. A. 2008. The great American schism: divergence of marine organisms after the rise of the Central American Isthmus. Annual Review of Ecology, Evolution, and Systematics 39: 63-91.

Levey, D. J. and F. G. Stiles. 1992. Evolutionary precursors of long-distance migration: resource availability and movement patterns in Neotropical landbirds. American Naturalist 140: 447-476.

Li, H. and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25: 1754-1760.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078-2079.

Lorenzen, E. D., R. Heller, and H. R. Siegismund. 2012. Comparative phylogeography of African savannah ungulates1. Molecular Ecology 21: 3656-3670.

Matocq, M. D., J. L. Patton, and M. N. F. da Silva. 2000. Population genetic structure of two ecologically distinct Amazonian spiny rats: separating history and current ecology. Evolution 54: 1423-1432.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and M. A. DePristo. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research 20: 1297-1303.

Melack, J. M. and L. L. Hess. 2011. Remote sensing of the distribution and extent of wetlands in the Amazon basin. Pages 43-59 in Amazonian Floodplain Forests. Springer, Netherlands.

Orme, D., R. Freckleton, G. Thomas, T. Petzoldt, S. Fritz, N. Isaac and W. Pearse. 2013. caper: Comparative Analyses of Phylogenetics and Evolution in R. R package version 0.5.2. Available from http://CRAN.R-project.org/package=caper (last accessed 5 July 2015).

Palumbi, S. R. 1992. Marine speciation on a small planet. Trends in Ecology and Evolution 7: 114-118.

Palumbi, S. R. 2003. Population genetics, demographic connectivity, and the design of marine reserves. Ecological Applications, 13: 146-158.

Parker III, T. A., D. F. Stotz, and J. W. Fitzpatrick. 1996. Ecological and distributional databases. Pages 113-436 in Neotropical Birds: Ecology and Conservation (D. F. Stotz, J. W. Fitzpatrick, T. A. Parker III, and D. Moskovits, eds.). University of Chicago Press, Chicago, Illinois.

Patton, J. L. and M. N. F. da Silva. 1998. Rivers, refuges, and ridges: the geography of speciation of Amazonian mammals. Pages 202-213 in Endless Forms: Species and Speciation (D. J. Howard and S. H. Berlocher, eds.).

Patton, J. L., M. N. F. da Silva, and J. R. Malcolm. 2000. Mammals of the Rio Juruá and the evolutionary and ecological diversification of Amazonia. Bulletin of the American Museum of Natural History 244: 1-306.

Paz, A., R. Ibáñez, K. R. Lips, and A. J. Crawford. 2015. Testing the role of ecology and life history in structuring genetic variation across a landscape: a trait-based phylogeographic approach. Molecular Ecology 24: 3723-3737.

Pearson, D. L. 1977. A pantropical comparison of bird community structure on six lowland forest sites. Condor 79: 232-244.

Prance, G. T. 1979. Notes on the vegetation of Amazonia III. The terminology of Amazonian forest types subject to inundation. Brittonia 31: 26-38.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155: 945-959.

R Core Team. 2015. R: A language and environment for statistical computing. Available from http://www.R-project.org/ (last accessed 20 October 2015).

Rambaut, A. and A. Drummond. 2007. Tracer v1.5. Available from http://beast.bio.ed.ac.uk/Tracer (last accessed 20 July 2015).

Reinhold, K. 1998. Sex linkage among genes controlling sexually selected traits. Behavioral Ecology and Sociobiology 44: 1-7.

Remsen, J. V., Jr., J. I. Areta, C. D. Cadena, A. Jaramillo, M. Nores, J. F. Pacheco, J. Pérez-Emán, M. B. Robbins, F. G. Stiles, D. F. Stotz, and K. J. Zimmer. A classification of the bird species of South America. American Ornithologists' Union. Available from http://www.museum.lsu.edu/~Remsen/SACCBaseline.html (last accessed 5 February 2015).

Remsen Jr, J. V. and T. A. Parker III. 1983. Contribution of river-created habitats to bird species richness in Amazonia. Biotropica 15: 223-231.

Robinson, S. K. and J. Terborgh 1995. Interspecific aggression and habitat selection by Amazonian birds. Journal of Animal Ecology 64: 1-11.

Ronquist, F., M. Teslenko, P. van der Mark, D. L. Ayres, A. Darling, S. Höhna, B. Larget, L. Liu, M. A. Suchard, and J. P. Huelsenbeck. 2012. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biology 61: 539-542.

Rosenberg, G. H. 1990. Habitat specialization and foraging behavior by birds of Amazonian river islands in northeastern Peru. Condor 92: 427-443.

Salisbury, C. L., N. Seddon, C. R. Cooney, and J. A. Tobias. 2012. The latitudinal gradient in dispersal constraints: ecological specialization drives diversification in tropical birds. Ecology Letters 15: 847-855.

Salo, J., R. Kalliola, I. Häkkinen, Y. Mäkinen, P. Niemelä, M. Puhakka, and P. D. Coley. 1986. River dynamics and the diversity of Amazon lowland forest. Nature 322: 254-258.

Schulenberg, T. S., D. F. Stotz, D. F. Lane, J. P. O'Neill, and T. A. Parker III. 2010. Birds of Peru: revised and updated edition. Princeton University Press, USA.

Smith, B. T., M. G. Harvey, B. C. Faircloth, T. C. Glenn, and R. T. Brumfield. 2014. Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. Systematic Biology 63: 83-95.

Smith, B. T., J. E. McCormack, A. M. Cuervo, M. J. Hickerson, A. Aleixo, C. D. Cadena, J. Pérez-Emán, C. W. Burney, X. Xie, M. G. Harvey, B. C. Faircloth, T. C. Glenn, E. P. Derryberry, J. Prejean, S. Fields, and R. T. Brumfield. 2014. The drivers of tropical speciation. Nature 515: 406-409.

Smith, J. V., E. L. Braun, and R. T. Kimball. 2013. Ratite non-monophyly: Independent evidence from 40 novel loci. Systematic Biology 62: 35-49.

Soltis, D. E., A. B. Morris, J. S. McLachlan, P. S. Manos, and P. S. Soltis. 2006. Comparative phylogeography of unglaciated eastern North America. Molecular Ecology 15: 4261-4293.

Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30: 1312-1313.

Stanley, S. M. 1975. A theory of evolution above the species level. Proceedings of the National Academy of Sciences of the U.S.A. 72: 646-650.

Sukumaran, J. and M. T. Holder. 2010. DendroPy: a Python library for phylogenetic computing. Bioinformatics 26: 1569-1571.

Taberlet, P., L. U. C. A. Fumagalli, A. G. Wust-Saucy, and J. F. Cosson. 1998. Comparative phylogeography and postglacial colonization routes in Europe. Molecular Ecology 7: 453-464.

Tajima, F. 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics, 105: 437-460.

Tuomisto, H. and K. Ruokolainen. 1997. The role of ecological knowledge in explaining biogeography and biodiversity in Amazonia. Biodiversity and Conservation 6: 347–357.

Wang, N., E. L. Braun, and R. T. Kimball. 2012. Testing hypotheses about the sister group of the Passeriformes using an independent 30-locus data set.Molecular biology and evolution, 29(2), 737-750.

Warren, W. C., D. F. Clayton, H. Ellegren, A. P. Arnold, L. W. Hillier, A. Künstner, S. Searle, S. White, A. J. Vilella, S. Fairley, A. Heger, L. Kong, C. P. Ponting, E. D. Jarvis, C. V. Mello, P. Minx, P. Lovell, T. A. F. Velho, M. Ferris, C. N. Balakrishnan, S. Sinha, C. Blatti, S. E. London, Y. Li, Y.-C. Lin, J. George, J. Sweedler, B. Southey, P. Gunaratne, M. Watson, K. Nam, N. Backström, L. Smeds, B. Nabholz, Y. Itoh, O. Whitney, A. P. Pfenning, J. Howard, M. Völker, B. M. Skinner, D. K. Griffin, L. Ye, W. M. McLaren, P. Flicek, V. Quesada, G. Velasco, C. Lopez-Otin, X. S. Puente, T. Olender, D. Lancet, A. F. A. Smit, R. Hubley, M. K. Konkel, J. A.Walker, M. A. Batzer, W. Gu, D. D. Pollock, L. Chen, Z. Cheng, E. E. Eichler, J. Stapley, J. Slate, R. Ekblom, T. Birkhead, T. Burke, D. Burt, C. Scharff, I. Adam, H. Richard, M. Sultan, A. Soldatov, H. Lehrach, S. V. Edwards, S.-P. Yang, X. Li, T. Graves, L. Fulton, J. Nelson, A. Chinwalla, S. Hou, E. R. Mardis, R. K. Wilson. 2010. The genome of a songbird. Nature 464: 757–762.

Watterson, G. A. 1975. On the number of segregating sites in genetical models without recombination. Theoretical Population Biology 7: 256-276.

Wittmann, F., W. J. Junk, and M. T. Piedade. 2004. The várzea forests in Amazonia: flooding and the highly dynamic geomorphology interact with natural forest succession. Forest Ecology and Management 196: 199-212.

Zerbino, D. R. and E. Birney. 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18: 821-829.

CHAPTER 5: POPULATION DIFFERENTIATION PREDICTS DIVERSIFICATION IN NEW WORLD BIRDS

INTRODUCTION

A grand challenge in evolutionary biology is understanding how population-level processes shape patterns of species diversity at deeper evolutionary time scales. Spatially structured, differentiated populations have long been viewed as potential incipient species (Mayr 1963), but it remains unclear if the rate at which differentiated populations form in a species is a limiting control on the diversification of its descendants (Allmon 1992, Barraclough and Nee 2001). Some have gone so far as to suggest that population differentiation is ephemeral and unimportant (Rosemblum et al. 2012), and recent studies have examined alternative macroevolutionary controls such as the evolutionary persistence of populations (Smith et al. 2014), the rate that reproductive isolation accrues between populations (Rabosky and Matute 2013), and the availability of ecological opportunities for new species (Price et al. 2014). However, some positive associations between taxonomic or indirect metrics of population differentiation and speciation rates have emerged from empirical datasets. For example, the number and rate of formation of taxonomic subspecies tends to predict the number of species or speciation rate in their group (Haskell and Adhikari 2009, Phillimore 2010), and factors thought to lead to the formation of population differentiation within species, such as dispersal ability, sometimes predict speciation rate (Jablonski 1986, Owens et al. 1999, Claramunt et al. 2012). The absence of standardized, quantitative estimates of population differentiation

and speciation rates from a large set of species has precluded a direct test of the link between population differentiation and diversification.

We assembled a population genetic dataset of mitochondrial gene sequences from 17,772 individuals representing a phylogenetically diverse group of 177 New World bird species (Fig. 1A, Appendix D). I defined species as all monophyletic, non-overlapping populations within a lineage. Any species lacking range-wide sampling was excluded. Species in the mitochondrial dataset occur in all biogeographic regions in the New World (Fig. 1B), and are representative of the breadth of ecological and life history variation found in New World birds. I estimated the number of genetically differentiated populations within each species using a Bayesian implementation of the Generalized Mixed Yule Coalescent model (bGMYC). The program clusters individuals by distinguishing coalescent processes within populations from Yule processes of diversification between populations (Pons et al. 2006, Reid and Carstens 2012) (Fig. 1C). The number of populations within species varied from zero to 35 with a mean of 4.49 (Fig. 2). To account for the variation in population number due to the age of the species, I estimated the rate at which new populations have formed since the crown age of the species (i.e., the age of the most recent common ancestor of extant haplotypes) using a constant rate pure-birth model (Magallón and Sanderson 2001). I estimated crown ages from time-calibrated phylogenies of each species estimated using a coalescent model and Bayesian inference (Drummond et al. 2012). The rate at which populations have split, hereafter the rate of population differentiation, varied from zero to 6.64 divergences per million years with an average of 0.78 divergences per million years (Fig. 2).

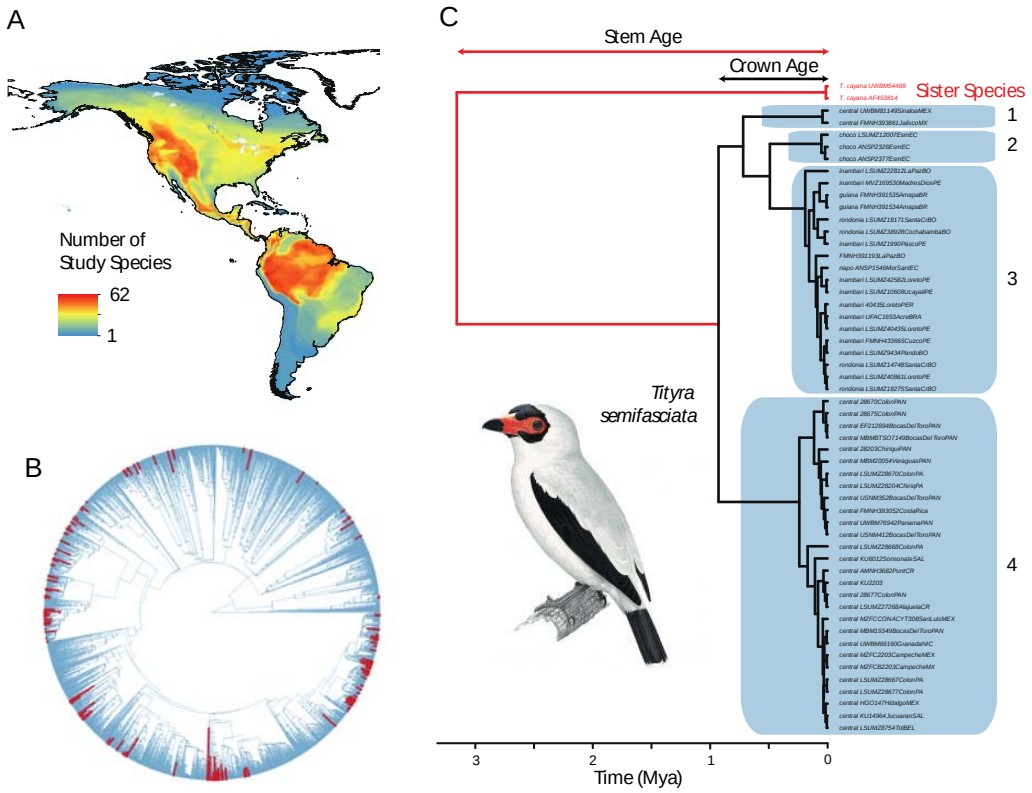


Figure 1. The sampling strategy and methodological approach used for this study. (A) Overlaid distribution maps from the New World bird species used to estimate amount of population genetic differentiation (n = 177). (B) The phylogenetic distribution of the study species within the tree of life of all birds (Jetz et al. 2012). The red branches indicate the species examined in this study and fall throughout the tree, presenting replicates of varying levels of phylogenetic independence for the purpose of comparative analysis. (C) An example of a mitochondrial gene tree used to estimate population divergence within one study species (*Tityra semifasciata*). The blue polygons represent population clusters for this species as inferred using bGMYC (Reid and Carstens 2012) based on a posterior probability threshold of shared population membership of 0.8. The stem age and crown age for this species, used to estimate rates of divergence, are also depicted.

We estimated speciation rates along the ancestral lineages leading to each of the 177 species in the population genetic datasets using a previously estimated phylogenetic tree of all bird species (Jetz et al. 2012). To account for taxa missing data in the tree, I analyzed a pruned version of the phylogeny that excluded all taxa lacking genetic data (33%) and incorporated an analytical correction to account for incomplete taxon

sampling (Rabosky et al. 2015). I estimated speciation and extinction rates using a Bayesian implementation of a model that jointly estimates 1) the number of distinct evolutionary rate regimes across a phylogenetic tree and 2) the speciation and extinction rates within each of the regimes (Rabosky 2014). The model assumes that the

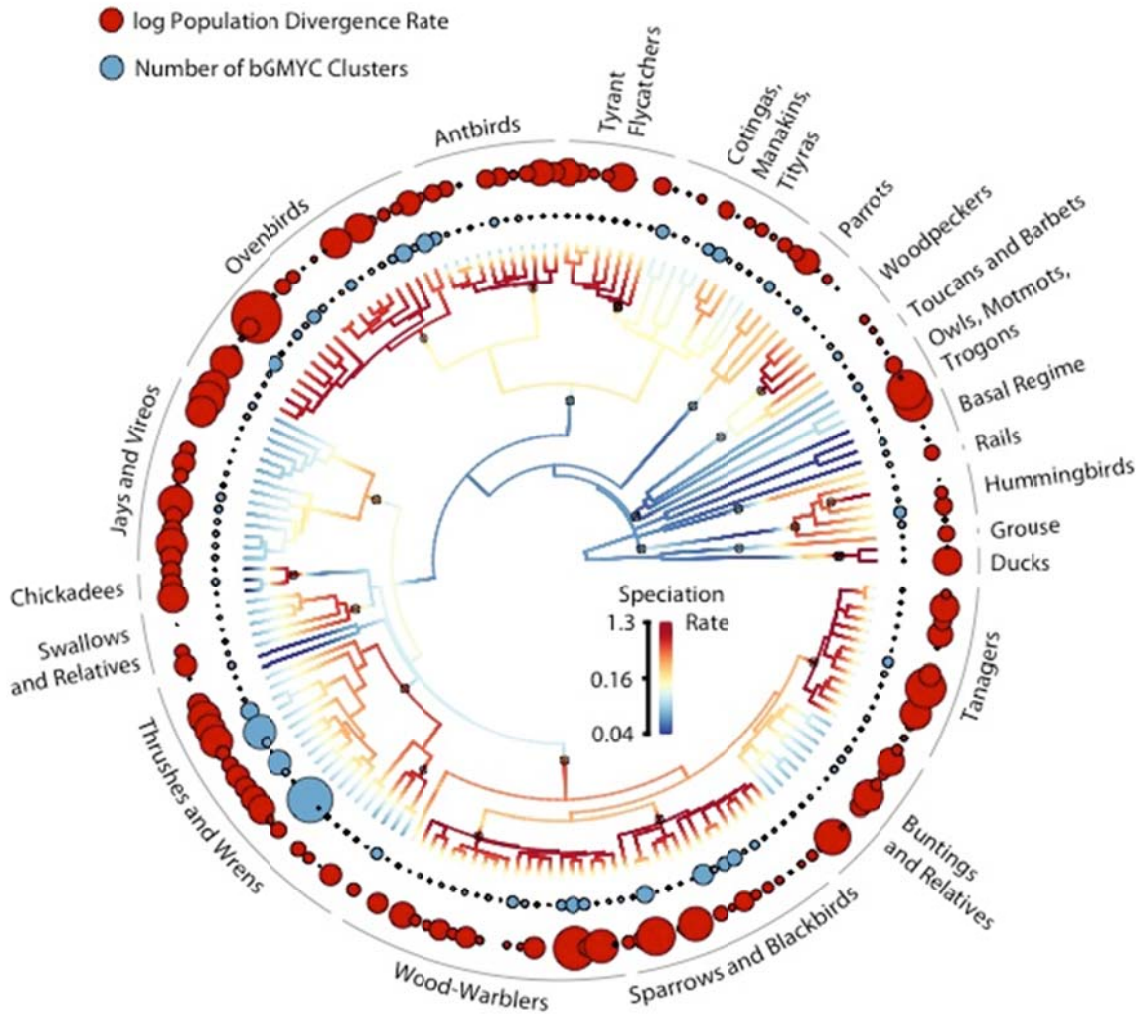


Figure 2. Circular phylogenetic tree of the 177 study species used to estimate rates of population differentiation rates, with a visual color spectrum depicting speciation rates along each branch. The significant speciation rate shifts (i.e. regime shifts) are indicated by black circles. Some branches include multiple circles, because the species used in speciation rate estimation but not represented in my population-level dataset are not shown. The diameter of the blue circles encircling the tree is proportional to the number of populations within the adjacent terminal species. The diameter of the red circles is proportional to the log-transformed rate of population differentiation.

phylogenetic tree was shaped by a heterogenous mixture of time-varying and constant-rate diversification processes (i.e. rate regimes). I found that bird diversification was best explained by 47 statistically distinguishable rate regimes, 22 of which included the 177 species in my population genetic dataset. The speciation rate across the 177 species varied from 0.04 to 0.73 species /My, with an average of 0.16 species/My (Fig. 2).

To test whether the rate of population genetic differentiation predicts the rate of new specialties formation, I used a trait-dependent diversification test that avoids phylogenetic pseudoreplication (Maddison and FitzJohn 2015) while accounting for autocorrelation in evolutionary rates across the branches of phylogenetic trees (Rabosky and Huang 2015). I found a significant positive correlation between population genetic differentiation rate and speciation rate (correlation coefficient (r) = 0.291, p = 0.004, Fig. 3). The correlation between population differentiation rate and speciation rate was significant for tropical species (r = 0.484, p = 0.002), but non-existent in species in the Temperate Zone. The difference in correlations between Tropical and Temperate was significant (p = 0.002) based on 1000 random permutations of species among latitudinal zones. The speciation rate averaged 5.57 times slower than the rate of population differentiation, suggesting that most populations fail to persist long enough to contribute to phylogenetic patterns at deeper time scales. This disparity was greater in the Temperate Zone (speciation rate averaged 6.63 times slower) than the Tropics (speciation rate averaged 4.80 times slower), although the difference was not significant based on 1000 random permutations. The correlation observed across all species was robust to the use of lower (0.7; r = 0.283, p = 0.004) and higher (0.9; r = 0.289, p = 0.006) posterior probability thresholds for assigning individuals to population clusters, to whether the

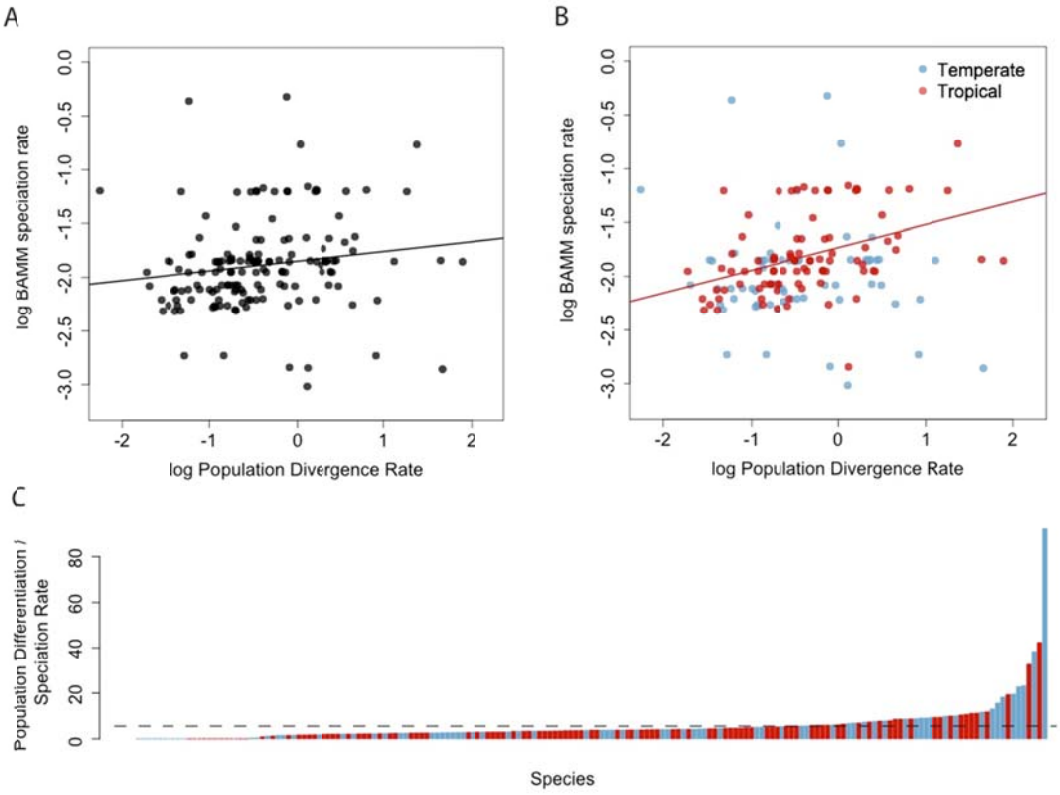


Figure 3. (A) Correlation between population differentiation rate and speciation rate in the 177 study species. (B) The same correlation as in (A) but with tropical and temperate species separated and a red trend line depicting the relationship between population differentiation rate and speciation rate for tropical species only. Trend lines are based on OLS regression, not the trait-dependent diversification test, and thus do not account for covariation among species from the same macroevolutionary regime. (C) The ratio of population differentiation to speciation rate for each study species. The dashed line depicts the mean value of 5.57. Blue bars indicate temperate species and red bars represent tropical species.

population differentiation rate was measured using the stem age rather than crown age of a species ($r = 0.332$, $p = 0.004$), to reductions in sampling in which 20% ($r = 0.274$, $p = 0.012$) and 40% ($r = 0.245$, $p = 0.016$) of individuals were pruned randomly within each species, and to models of population differentiation incorporating moderate ($eps = 0.45$; $r = 0.291$, $p = 0.006$) and high ($eps = 0.9$; $r = 0.265$, $p = 0.014$) levels of extinction (see also SI). Using simulations I also found that the Type 1 error rate, which is problematic in

many tests of trait-dependent diversification (Rabosky and Huang 2015), was low (0.098) given the covariance structure of my data. The correlation between the population genetic differentiation rate and the speciation rate was robust to the taxonomy used to circumscribe species for the population-level analysis, with a more finely subdivided taxonomy producing similar results to the primary taxonomy I examined ($r = 0.210$, $p = 0.014$; Tropical: $r = 0.416$, $p = 0.010$; Temperate: NS). I found no correlation between the raw number of population clusters and the speciation rate, suggesting that the rate of population genetic differentiation, but not the level of standing differentiation, predicts the rate of speciation.

The correlation between population differentiation rate and speciation rate suggests that speciation is an important control on diversification. However, the probability that differentiated populations will persist into deeper time, rather than the rate at which they form and differentiate, could also influence the speciation rate (Mayr 1963, Stanley 1979, Etienne et al. 2014). To examine this, I measured in all 177 species the length of the stem branch, the time between when a species diverged from its sister group and the oldest divergence event between extant populations, relative to the stem age. Stem branch length can be used as an index of the prevalence of population extinction in each lineage (Nee et al. 1994). This 'extinction index' is the inverse of persistence. I found no correlation between the extinction index and the speciation rate, either in the entire data set or in the tropical or temperate species treated separately. my data thus suggest that population persistence does not predict the speciation rate, or at least has much less predictive power than population differentiation rate.

Another potential control on speciation rate is the rate at which reproductive isolation forms between incipient species (Dobzhansky 1937, Mayr 1942). A recent comparative study using the same avian phylogenetic dataset examined here found no association between the rate at which intrinsic postzygotic reproductive isolation arises and speciation rate (Rabosky and Matute 2013). The time to loss of intrinsic hybrid fertility and viability may be long enough, at least in birds, that intrinsic post-mating reproductive isolation does not exert a strong control on diversification (Price and Bouvier 2002). The rate of formation of pre-mating isolation is an important alternative that I were unable to address (Price 2008).

Ecological limits, or constraints on species diversity due to total resource availability, may erode the relationship between the rate of population differentiation and the rate of speciation (Valentine and Moores 1972, Rosenzweig 1975, Price et al. 2014). A slowdown in speciation rate in a clade may be evidence of diversity dependence due to the action of ecological limits. The 2.6-fold average slowdown in speciation rate observed across the 22 evolutionary regimes in the avian phylogeny is consistent with the action of such limits. If ecological limits reduce the association between population differentiation and speciation rate, I might expect the correlation to be weaker in clades with a signal of diversity dependence. I divided the dataset into clades with greater slowdowns and those with weak or no slowdowns, and failed to find a stronger relationship between population differentiation rate and speciation rate in those with weak or no slowdowns. However, "early burst" patterns in phylogenies may result from biased taxon sampling or the use of nucleotide substitution models that fail to capture molecular evolutionary processes (Moen and Morlon 2014), and rate shifts in clades

characterized by diversity dependence may degrade after extended periods at equilibrium (Rabosky and Hurlbert 2015).

We found that population genetic differentiation predicts speciation rate, despite the existence of competing potential controls on speciation, such as population persistence, the rate of evolution of reproductive isolation, and ecological limits to diversity. Further investigation of this association, however, is warranted. Given that lineages with greater rates of speciation necessarily do not contain old stem and crown ages, which in turn results in fewer lineages with slow population differentiation rates, it might be suggested that the relationship I recovered is a statistical artifact of the non-independence of estimates of speciation and differentiation rates. However, the boundary between statistical artifact and biological relationship in this case is not clear, since the constraint on crown and stem age is also a biological phenomenon resulting from short waiting times to speciation. Regardless, my results support the importance of continued investigation of population-level processes as potential drivers of the evolution of diversity. If traits exist that predispose species to population divergence, they may be responsible for long-term differences in speciation or extinction rates across lineages, a process known as species selection (Stanley 1975). I anticipate more and larger datasets will provide more evidence for the important role of microevolutionary processes such as divergence, persistence, natural and sexual selection, and hybridization on the macroevolutionary dynamics that have produced the remarkable diversity of organisms worldwide.

REFERENCES

Allmon, W. D. 1992. A causal analysis of stages in allopatric speciation. Oxford Surveys in Evolutionary Biology 8: 219-257.

Barraclough, T. G. and S. Nee. 2001. Phylogenetics and speciation. Trends in Ecology and Evolution 16: 391-399.

Claramunt, S., E. P. Derryberry, J. V. Remsen Jr., and R. T. Brumfield. 2012. High dispersal ability inhibits speciation in a continental radiation of passerine birds. Proceedings of the Royal Society of London B 279: 1567-1574.

Drummond, A. J., M. A. Suchard, D. Xie, and A. Rambaut. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molecular Biology and Evolution 29: 1969-1973.

Etienne, R. S., H. Morlon, and A. Lambert. 2014. Estimating the duration of speciation. Evolution 68: 2430-2440.

Haskell, D. G. and A. Adhikari. 2009. Darwin's manufactory hypothesis is confirmed and predicts extinction risk of extant birds. PLoS One 4: e5460.

Jablonski, D. 1986. Larval ecology and macroevolution in marine invertebrates. Bulletin of Marine Science 39: 565-587.

Jetz, W., G. H. Thomas, J. B. Joy, K. Hartmann, and A. O. Mooers. 2012. The global diversity of birds in space and time. Nature 491: 444-448.

Maddison, W. P. and R. G. FitzJohn. 2015. The unsolved challenge to phylogenetic correlation tests for categorical characters. Systematic Biology 64: 127-136.

Magallón, S. and M. J. Sanderson. 2001. Absolute diversification rates in angiosperm clades. Evolution 55: 1762-1780.

Mayr, E. 1963. Animal Species and Evolution. Belknap, Cambridge, MA.

Nee, S., E. C. Holmes, R. M. May, and P. H. Harvey. 1994. Extinction rates can be estimated from molecular phylogenies. Philosophical Transactions of the Royal Society of London B 344: 77-82.

Phillimore, A. B. 2010. Subspecies origination and extinction in birds. Ornithological Monographs 67: 42-53.

Owens, I. P. F., P. M. Bennett, and P. H. Harvey. 1999. Species richness among birds: body size, life history, sexual selection, or ecology? Proceedings of the Royal Society of London B 266: 933-939.

Pons, J., T. G. Barraclough, J. Gomez-Zurita, A. Cardoso, D. P. Duran, S. Hazell, S. Kamoun, W. D. Sumlin, and A. P. Vogler. 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. Systematic Biology 55: 595-609.

Price, T. D., D. M. Hooper, C. D. Buchanan, U. S. Johansson, D. T. Tietze, P. Alström, U. Olsson, M. Ghosh-Harihar, F. Ishtiaq, S. K. Gupta, J. Martens, B. Harr, P. Singh, and D. Mohan. 2014. Niche filling slows the diversification of Himalayan songbirds. Nature 509: 222-225.

Rabosky, D. L. 2014. Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. PLoS One 9: e89543.

Rabosky, D. L. and H. Huang. 2015. A robust semi-parametric test for detecting trait-dependent diversification. Systematic Biology doi:10.1093/sysbio/syv066.

Rabosky, D. L. and D. R. Matute. 2013. Macroevolutionary speciation rates are decoupled from the evolution of intrinsic reproductive isolation in Drosophila and birds. Proc. Nat. Acad. USA 110, 15354-15359.

Rabosky, D. L., P. O. Title, and H. Huang. 2015. Minimal effects of latitude on present-day speciation rates in New World birds. Proceedings of the Royal Society of London B 282: 20142889.

Reid, N. M. and B. C. Carstens. 2012. Phylogenetic estimation error can decrease the accuracy of species delimitation: a Bayesian implementation of the general mixed Yule-coalescent model. BMC Evolutionary Biology 12: 196.

Rosenblum, E. B., B. A. J. Sarver, J. W. Brown, S. Des Roches, K. M. Hardwick, T. D. Hether, J. M. Eastman, M. W. Pennell, and L. J. Harmon. 2012. Goldilocks meets Santa Rosalia: An ephemeral speciation model explains patterns of diversification across time scales. Evolutionary Biology 39: 255-261.

Smith, B. T., J. E. McCormack, A. M. Cuervo, M. J. Hickerson, A. Aleixo, C. D. Cadena, J. Pérez-Emán, C. W. Burney, X. Xie, M. G. Harvey, B. C. Faircloth, T. C. Glenn, E. P. Derryberry, J. Prejean, S. Fields, and R. T. Brumfield. 2014. The drivers of tropical speciation. Nature 515: 406-409.

Stanley, S. M. 1979. Macroevolution: Pattern and Process. W. H. Freeman, San Francisco, CA.

Dobzhansky, T. 1937. Genetics and the Origin of Species. Columbia University Press, New York.

Mayr, E. 1942. Systematics and the Origin of Species from the Viewpoint of a Zoologist. Columbia University Press, New York.

Schluter, D. 2009. Evidence for ecological speciation and its alternative. Science 323: 737-741.

Price, T. D. and M. M. Bouvier. 2002. The evolution of F1 postzygotic incompatibilities in birds. Evolution 56: 2083-2089.

Price, T. D. 2008. Speciation in Birds. Roberts and Co., Greenwood Village, CO.

Valentine, J. W. and E. M. Moores, 1972. Global tectonics and the fossil record. Journal of Geology 80: 167-184.

Rosenzweig, M. L. 1975. On continental steady states of species diversity. Pages 121-140 in Ecology and Evolution of Communities (M. L. Cody, J. M. Diamond, eds.). Belknap, Cambridge, MA.

Moen, D. and H. Morlon. 2014. Why does diversification slow down? Trends in Ecology and Evolution 29: 190-197.

Rabosky, D. L. and A. H. Hurlbert. 2015. Species richness at continental scales is dominated by ecological limits. American Naturalist 185: 572-583.

Stanley, S. M. 1975. A theory of evolution above the species level. Proceedings of the National Academy of Sciences of the U.S.A. 72: 646-650.

CHAPTER 6: CONCLUSIONS

Understanding the processes responsible for geographic variation and its significance has long been a major goal of evolutionary biology (Gould and Johnston 1972). Differences in variation across geographic areas, species, or time periods, analyzed in a comparative framework, can help us understand the causes of geographic variation. Most comparative studies of geographic variation, however, have been limited by the number of species available for comparison or by the resolution of estimates of geographic diversity and population history. In this dissertation, I tested and used strategies for obtaining genomic data to obtain more detailed estimates of geographic variation and population history and compiled large (both in number of species and number of independent markers) comparative datasets in order to evaluate the causes and significance of geographic variation.

In Chapter 2, I compared two promising methods for obtaining large, genome-wide datasets from which to estimate population history: restriction site associated DNA sequencing (RAD-Seq) and sequence capture of ultraconserved elements (UCEs). I found that both methods were useful for different applications. In particular, RAD-Seq is more appropriate for obtaining large numbers of loci from single species with shallow population histories, whereas sequence capture may be better suited to comparative studies involving multiple species. These results directed strategies for my subsequent chapters, and should also serve as a useful guide for other researchers studying population genomics in non-model species.

In Chapter 3, I used RAD-Seq to study the phylogeography of a single widespread Neotropical bird species (*Xenops minutus*). I found that geographic variation was a result of structuring of populations by landscape barriers and also found evidence of historical population expansion and migration between populations from different areas. These results indicated that genome-wide data could be useful for examining diverse historical processes that might be responsible for geographic variation.

Chapter 4 involved applying sequence capture of UCEs and exons to 20 pairs of species or species complexes that included a species distributed in floodplain forest and a close relative distributed in upland forest in the Amazon. Comparisons across species revealed that the ecology of a species impacted population genomic diversity and history. Upland forest species had higher genetic diversity, greater divergence between populations, and deeper histories than floodplain forest species. This result provided the first evidence from a large sample of species that habitat association predicts population genetics in birds, and also demonstrated that genomic data can be used to estimate genetic diversity and population histories for comparative studies across many species.

I sought to determine the significance of differences across species in levels of geographic variation in my last research chapter. Although intuition would suggest that species containing greater geographic divergence would exhibit higher speciation rates through time, this connection has not been adequately demonstrated (Allmon 1992, Barraclough and Nee 2001). I compiled mitochondrial phylogeography datasets from 177 New World bird species to compare metrics of population divergence with speciation rates estimated from existing phylogenetic estimates of all birds. Population divergence rate predicted speciation rate in the ancestral lineage of a species, a relationship that was

strongest in the Tropics. I also found that population divergence occurred at nearly six times the rate of speciation, on average, suggesting that even though population divergence predicts speciation, many divergent populations do not persist to form species. This study provides the first demonstration that quantitative estimates of geographic variation within species are associated with speciation rates, which implies that traits that promote population divergence may have impacts on macroevolutionary diversity over the course of organismal diversification.

The sources of geographic variation are diverse and not mutually exclusive (Gould and Johnston 1972, Antonelli et al. 2010). Future studies sampling more species and examining better models of processes that shape genome-wide genetic diversity promise to add greatly to my knowledge of which processes my important and their relative contributions to variation. My dissertation is one of the first studies to examine geographic variation across species using genomic datasets and represents an important first step in this direction.

Tying the chapters together, traits that predict population divergence within species may result in higher speciation rates over long evolutionary timescales in those lineages. Bird lineages that inhabit upland forest, for example, may have higher rates of speciation that those that inhabit floodplain forest because the species within them experience higher rates of population divergence. Repeated across a diversity of lineages, this effect might lead to the proliferation of traits that promote population divergence, and effect known as species selection (Stanley 1975). Such an effect could lead to a higher diversity of upland forest bird species over time, for example, than floodplain forest species. Interestingly, the upland forest avifauna is more diverse (1,058 species) than the

floodplain forest avifauna (154 species), and it is possible species selection between habitats has played a role in this disparity. Genomic datasets combined with comparative analyses may allow us to address this fundamental question in the near future.

REFERENCES

Allmon, W. D. 1992. A causal analysis of stages in allopatric speciation. Oxford Surveys in Evolutionary Biology 8: 219-257.

Antonelli, A., A. Quijada-Mascareñas, A. J. Crawford, J. M. Bates, P. M. Velasco, W. Wüster. 2010. Molecular studies and phylogeography of Amazonian tetrapods and their relation to geological and climatic models. Pages 386-404 in Amazonia, Landscape and Species Evolution: A Look into the Past (C. Hoorn and F. P. Wesselingh, eds.). Wiley-Blackwell, London, UK.

Barraclough, T. G. and S. Nee. 2001. Phylogenetics and speciation. Trends in Ecology and Evolution 16: 391-399.

Gould, S. J. and R. F. Johnston. 1972. Geographic variation. Annual Review of Ecology and Systematics 3: 457-498.

Stanley, S. M. 1975. A theory of evolution above the species level. Proceedings of the National Academy of Sciences of the U.S.A. 72: 646-650.

APPENDIX A: SUPPLEMENTARY MATERIAL TO CHAPTER 2

Sample List

	Museum[a]	Tissue #	Biogeog. Area	Subspecies	Country	State	Locality	Lat.	Long.
1	KUMNH	2044	C. America	*X. m. mexicanus*	Mexico	Campeche	Calakmul, El Arroyo, 6 km S Silvituc	18.5928	-90.2561
2	LSUMZ	60935	C. America	*X. m. mexicanus*	Honduras	Cortés	Cerro Azul Meamber National Park, Los Pinos	14.8728	-87.9050
3	LSUMZ	2209	Chocó	*X. m. littoralis*	Panama	Darién	Cana on E slope Cerro Pirré	7.7560	-77.6840
4	LSUMZ	11948	Chocó	*X. m. littoralis*	Ecuador	Esmeraldas	El Placer	0.8667	-78.5500
5	LSUMZ	4244	Napo	*X. m. obsoletus*	Peru	Loreto	Lower Rio Napo, E bank Rio Yanayacu, ca. 90 km N Iquitos	-2.8200	-73.2738
6	LSUMZ	6862	Napo	*X. m. obsoletus*	Peru	Loreto	5 km N Amazon River, 85 km NE Iquitos	-3.4167	-72.5833
7	LSUMZ	9026	Inambari	*X. m. obsoletus*	Bolivia	Pando	Nicolás Suarez, 12 km by road S Cobija, 8 km W on road to Mucden	-11.4703	-68.7786
8	FMNH	433364	Inambari	*X. m. obsoletus*	Peru	Cusco	Consuelo, 15.9 km SW Pilcopata	-13.0167	-71.4833

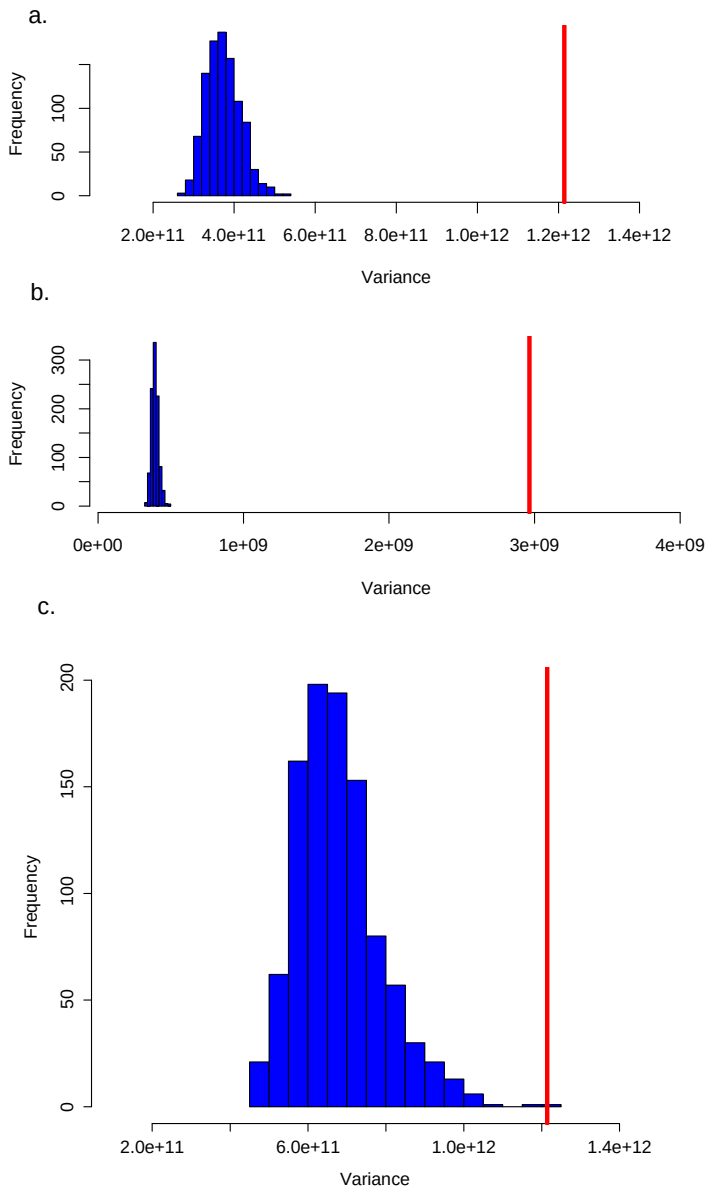
[a] Museums are University of Kansas Natural History Museum (KUNHM), Field Museum (FMNH), and Louisiana State University Museum of Natural Science (LSUMZ).

APPENDIX A CONT. Summary Statistics

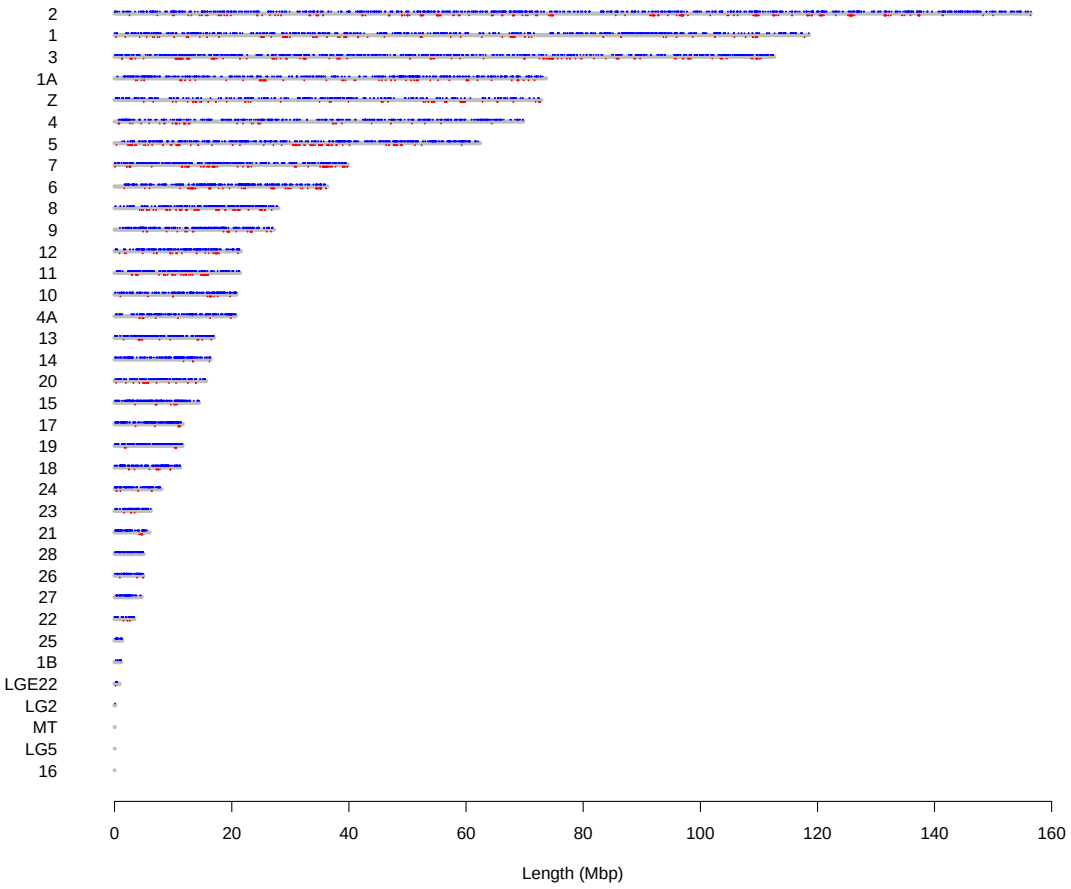
	Raw Reads		Reads in Assembly (96% Similarity, 7x read depth)	
Sample	Sequence Capture	RAD-Seq	Sequence Capture	RAD-Seq
LSUMZ 11948	15414860	2373918	13898684	1081825
FMNH 433364	9853008	3012717	8884732	1454397
LSUMZ 6862	9027970	2989976	8129564	1403940
KUMNH 2044	12213310	1386861	10913706	390501
LSUMZ 2209	14538892	2623628	13095518	1198260
LZUMZ 9026	10624014	3178928	9630866	1466441
LSUMZ 4244	12308350	3108856	11113426	1469613
LSUMZ 60935	12100066	2311639	10939708	1037481

APPENDIX A CONT. BPP Results from two runs from RAD-Seq and UCE data

Summary Statistic	mean	stderr of mean	median	geometric mean	95% HPD lower	95% HPD upper	auto-correlation time (ACT)	effective sample size (ESS)
GBSa thetaA	1.49E-03	2.78E-06	1.49E-03	1.49E-03	1.30E-03	1.69E-03	554.8756	1297.5918
GBSa thetaB	3.45E-03	5.15E-06	3.45E-03	3.44E-03	3.05E-03	3.85E-03	456.4912	1577.2527
GBSa thetaAB	6.35E-03	8.39E-06	6.35E-03	6.35E-03	5.74E-03	6.95E-03	531.1073	1355.6621
GBSa tauAB	1.41E-03	3.88E-06	1.41E-03	1.40E-03	1.20E-03	1.63E-03	937.3742	768.1052
GBSb thetaA	1.48E-03	2.66E-06	1.48E-03	1.48E-03	1.29E-03	1.68E-03	522.5401	1377.8885
GBSb thetaB	3.62E-03	4.83E-06	3.62E-03	3.62E-03	3.21E-03	4.04E-03	376.1034	1914.3727
GBSb thetaAB	6.36E-03	7.91E-06	6.36E-03	6.36E-03	5.77E-03	6.97E-03	473.5206	1520.5294
GBSb tauAB	1.45E-03	3.75E-06	1.45E-03	1.45E-03	1.24E-03	1.66E-03	882.5502	815.8199
UCEa thetaA	2.01E-03	6.78E-07	2.00E-03	2.00E-03	1.81E-03	2.21E-03	31.0538	23185.5981
UCEa thetaB	3.66E-03	1.64E-06	3.65E-03	3.65E-03	3.24E-03	4.07E-03	42.9481	16764.4708
UCEa thetaAB	1.40E-03	1.64E-06	1.39E-03	1.39E-03	1.15E-03	1.63E-03	131.4151	5478.837
UCEa tauAB	1.30E-03	7.84E-07	1.30E-03	1.29E-03	1.19E-03	1.41E-03	137.7731	5225.9978
UCEb thetaA	2.19E-03	7.50E-07	2.19E-03	2.19E-03	1.97E-03	2.43E-03	29.5317	24380.6551
UCEb thetaB	4.10E-03	1.84E-06	4.09E-03	4.10E-03	3.63E-03	4.59E-03	40.3285	17853.4106
UCEb thetaAB	2.02E-03	1.40E-06	2.02E-03	2.01E-03	1.75E-03	2.29E-03	74.0813	9719.0728
UCEb tauAB	1.30E-03	7.17E-07	1.30E-03	1.30E-03	1.19E-03	1.41E-03	114.3952	6293.9895

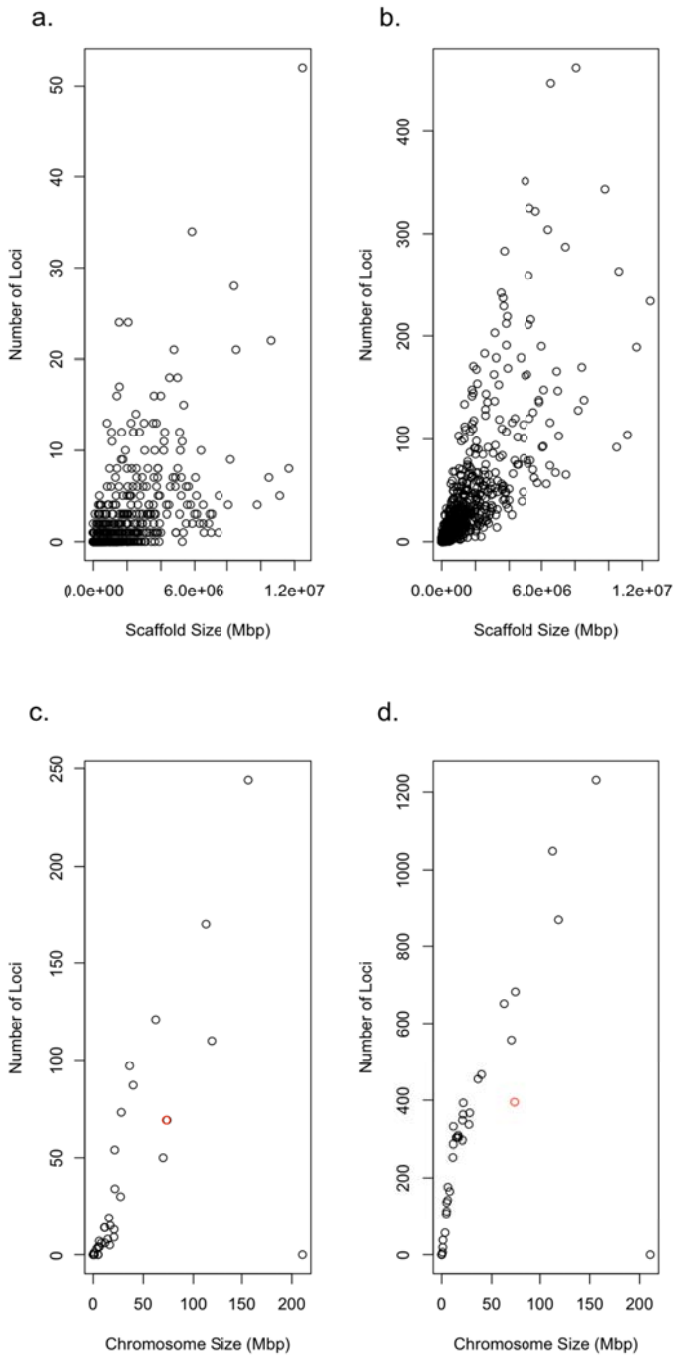


Assessment of clustering of loci relative to simulated random distributions. **(a)** The variance among *Manacus vitellinus* scaffolds in mean distance between loci from sequence capture (red line) compared to the values from 1000 simulations in which sequence capture loci are plotted randomly across scaffolds contingent on their length (blue bars). **(b)** The variance among scaffolds in mean distance between loci for RAD-Seq loci (red line) compared to the values from 1000 simulations in which RAD-Seq loci are plotted randomly across scaffolds contingent on their length (blue bars). **(c)** The variance among scaffolds in mean distance between loci from sequence capture (red line) relative to random subsets of the RAD-Seq loci of the same size as the sequence capture dataset.

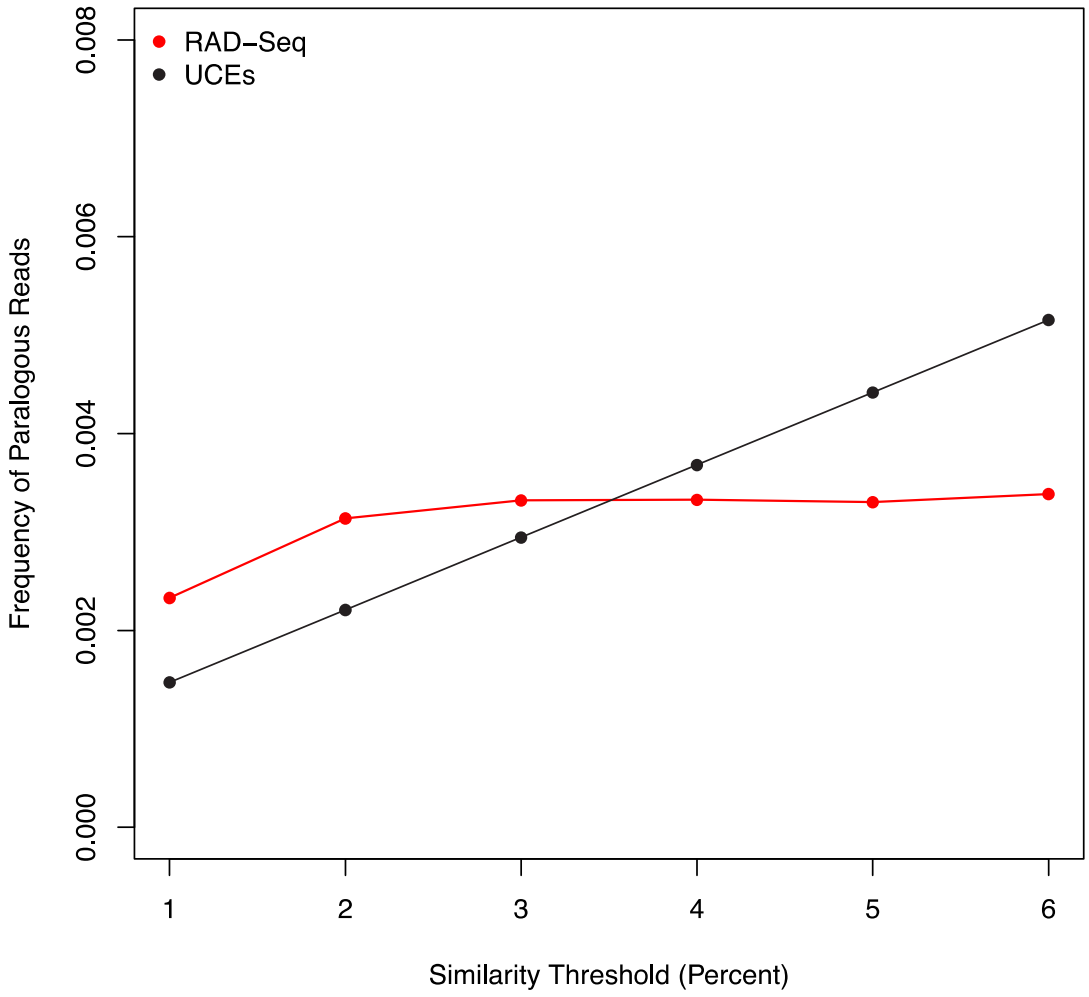


The distribution of RAD-Seq loci (blue dots) and sequence capture loci (red dots) across *Taeniopygia guttata* chromosomes based on Blastn mapping results.

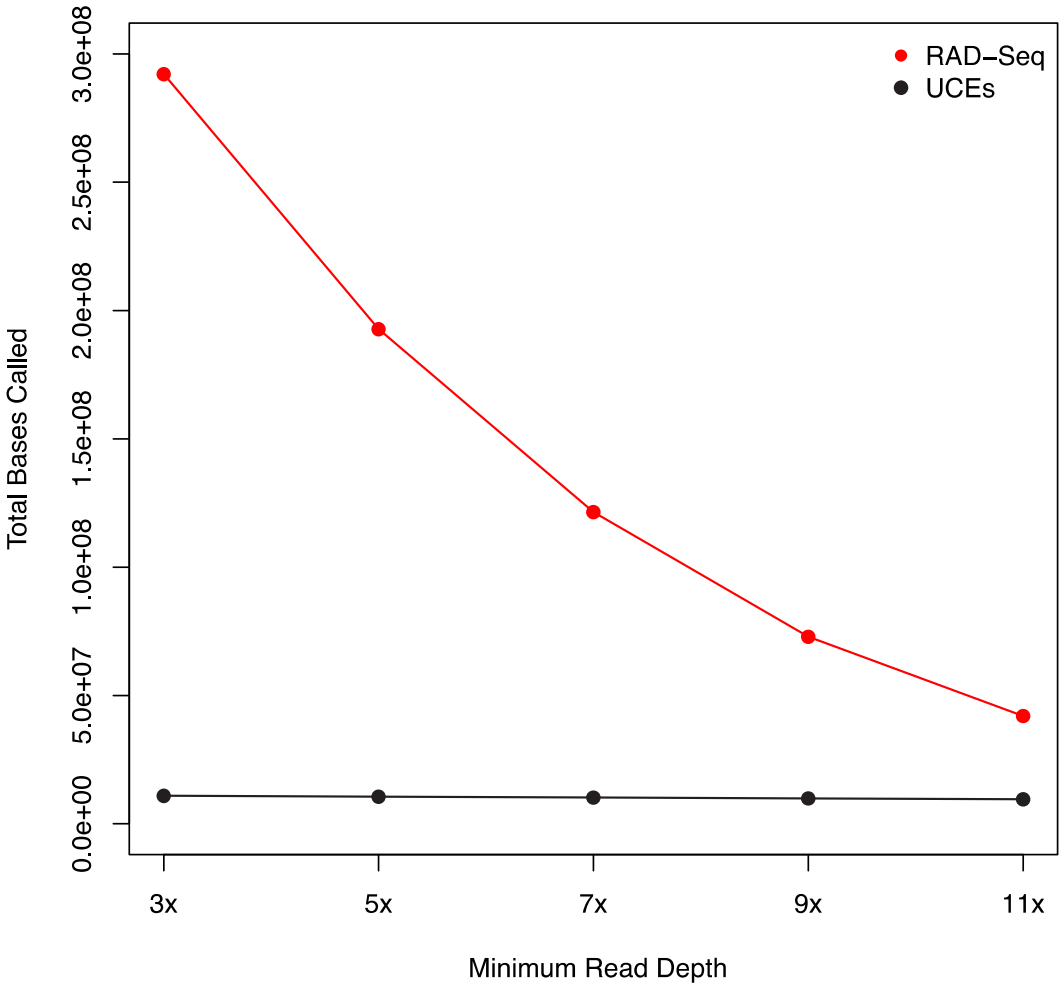
APPENDIX A CONT.



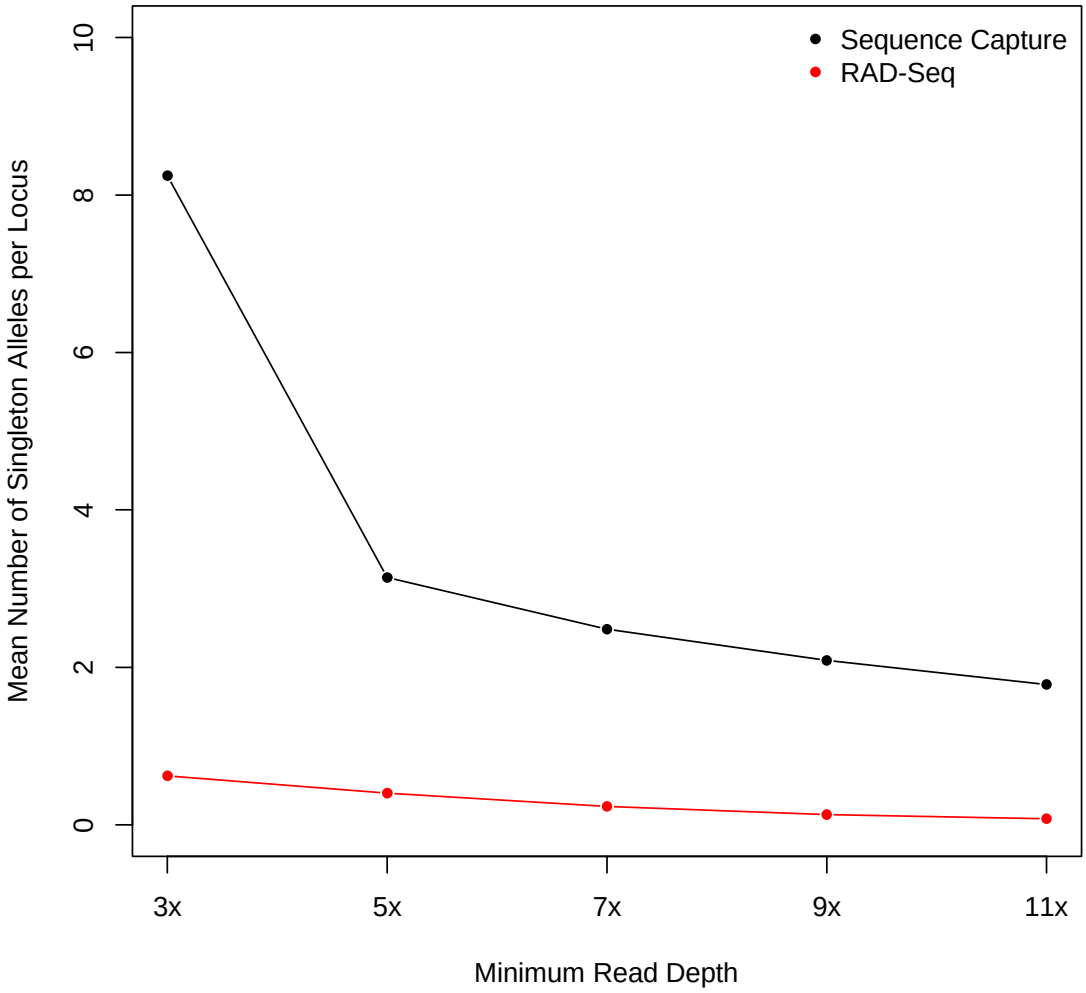
The number of loci relative to scaffold size in the *Manacus vitellinus* genome for (a) sequence capture loci and (b) RAD-Seq loci, and relative to chromosome size in the *Taeniopygia guttata* genome for (c) sequence capture loci and (d) RAD-Seq loci. In (c) and (d) the point representing the sex-linked Z chromosome is colored red.



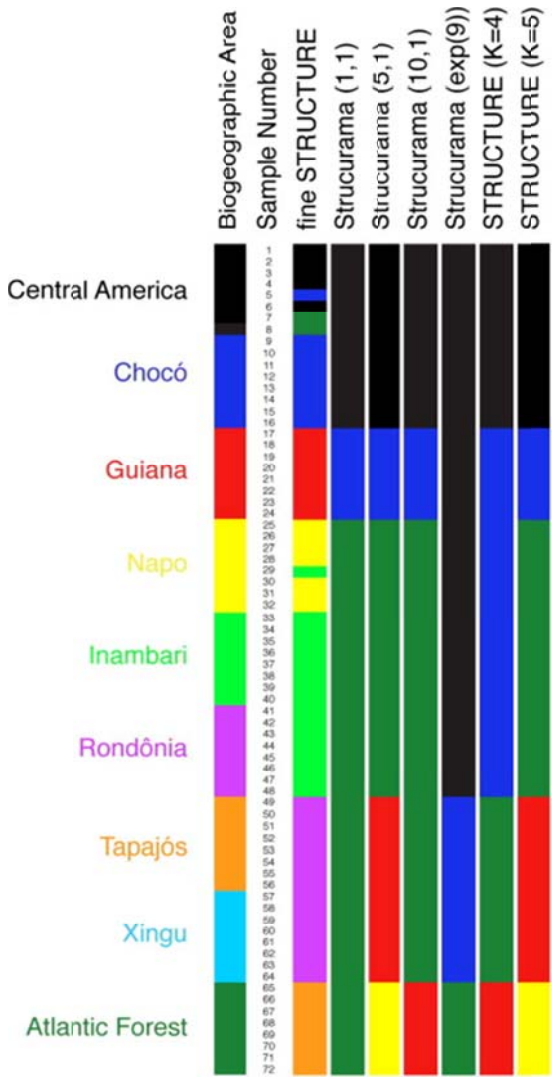
The impact of the similarity threshold (here in % divergence allowed between reads for assembly) on the absolute frequency of loci containing paralogous reads (under-splitting).



The impact of the minimum read depth used to call alleles on the total bases in final alignments.

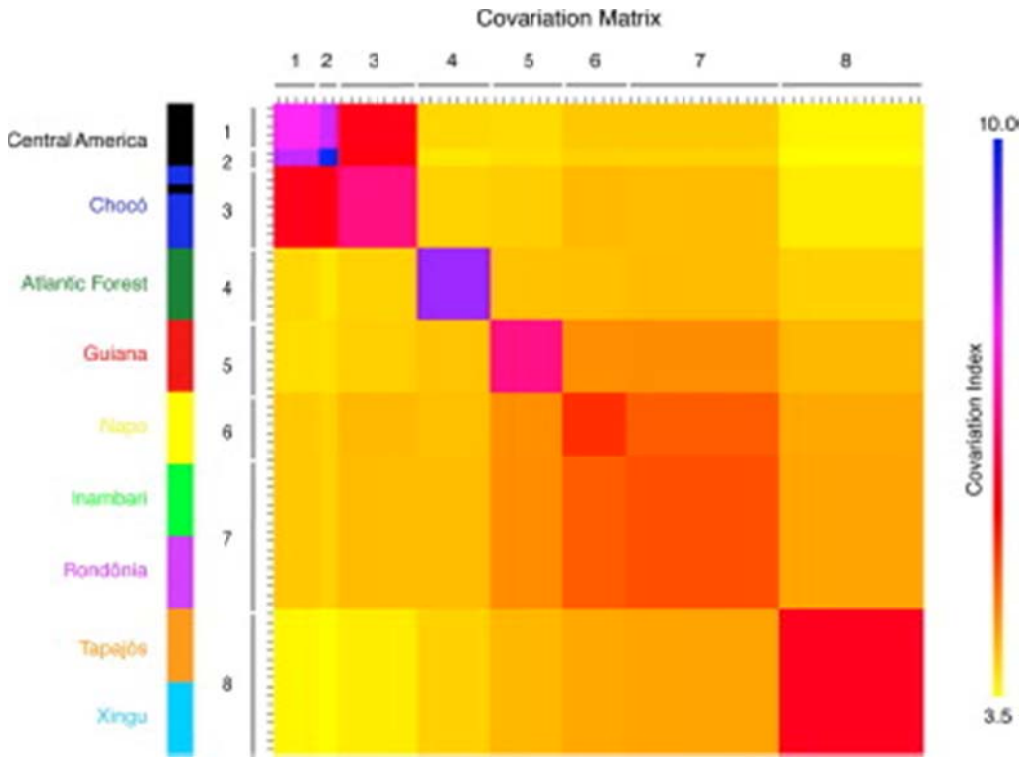


The impact of the minimum read depth used to call alleles on the number of singleton alleles recovered per locus.



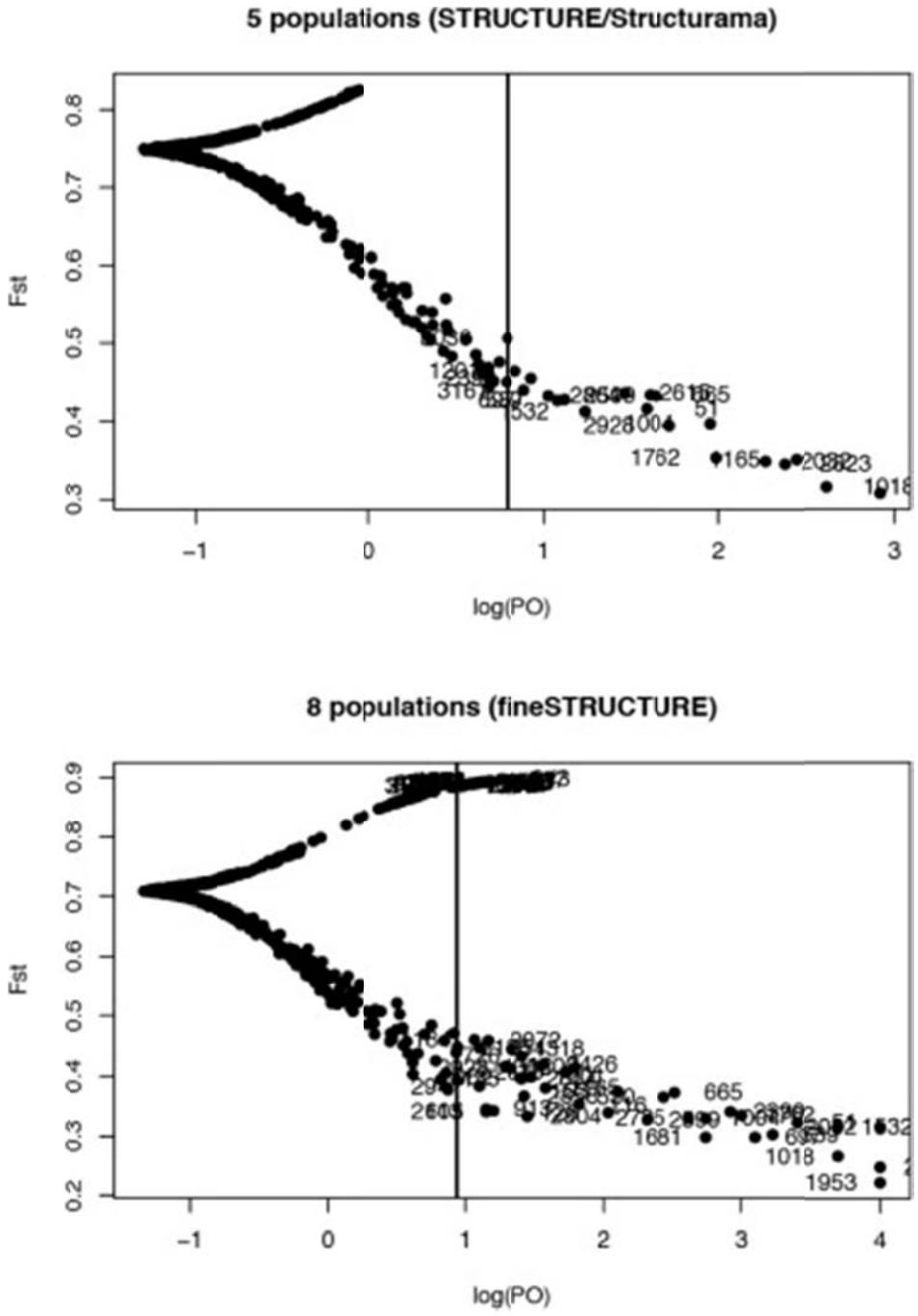
Numbers of populations and population assignments from a subset of the clustering analyses conducted in fineSTRUCTURE, STRUCTURE, and Structurama. Colors in the population assignment columns distinguish populations, but are not necessarily related between columns, nor do they refer to biogeographic areas

APPENDIX B CONT.



The covariance matrix from the fineSTRUCTURE analysis showing populations identified by fineSTRUCTURE and the geographic area associated with each individual. Higher values in the covariation index correspond to greater similarity between individuals.

APPENDIX B CONT.



Plots of F_{st} for all loci from the BayeScan outlier analysis of both the STRUCTURE/ Structurama populations and fineSTRUCTURE populations show positive (fineSTRUCTURE) and negative (both analyses) outliers that exceed the posterior odds ratio set based on an expected false discovery rate of 0.05.

APPENDIX B CONT. Sample information for all individuals used in this study. The number column refers to sample numbers referenced elsewhere in the paper. Museum abbreviations correspond to: (ANSP) Academy of Natural Sciences of Drexel University, Philadelphia, USA; (CUMV) Cornell University Museum of Vertebrates, Ithaca, NY, USA; (KUMNH) Kansas University Museum of Natural History, Lawrence, KS, USA; (LSUMZ) Louisiana State University Museum of Natural Science, Baton Rouge, LA, USA; (MZFC) Museo de Zoología "Alfonso L. Herrera" de la Facultad de Ciencias, Universidad Nacional Autónoma de México, DF, México; (MBM) Marjorie Barrick Museum, University of Nevada - Las Vegas, NV, USA now housed at the Burke Museum, University of Washington, Seattle, USA; (MPEG) Museu Paraense Emílio Goeldi, Belém, Brasil; (MZUSP) Museu de Zoologia da Universidade de São Paulo, Brasil; and (USNM) National Museum of Natural History, Smithsonian Institution, Washington, DC.

#	Museum	Tissue Number	Biog. Area	Subspecies	Country	State	Locality	Lat.	Long.
1	LSUMZ	35767	Central America	*X. m. ridgwayi*	Costa Rica	Cartago	11 km SW Pejibaye	9.7833	-83.7500
2	LSUMZ	60935	Central America	*X. m. mexicanus*	Honduras	Cortés	Cerro Azul Meamber National Park, Los Pinos	14.8728	-87.9050
3	USNM	1283	Central America	*X. m. ridgwayi*	Panama	Bocas Del Toro	Valiente Peninsula, Punta Alegre, N. Bahia Azul	9.0215	-81.7620
4	CUMV	50919	Central America	*X. m. ridgwayi*	Panama	Chiriquí	Burica Peninsula, 100-160m	8.0333	-82.8667
5	CUMV	50738	Central America	*X. m. ridgwayi*	Panama	Coclé	El Cope National Park	8.6698	-80.5930
6	KUMNH	2044	Central America	*X. m. mexicanus*	Mexico	Campeche	Calakmul, El Arroyo, 6 km S Silvituc	18.5928	-90.2561
7	MZFC	51	Central America	*X. m. mexicanus*	Mexico	Chiapas	N portion of La Omega, Monumento Natural Yaxchilan	16.9017	-90.9733
8	MZFC	238	Central America	*X. m. mexicanus*	Mexico	Oaxaca	20 km NE Chalchijapa	17.0667	-94.5833
9	ANSP	2227	Chocó	*X. m. littoralis*	Ecuador	Esmeraldas	20 km NNW Alto Tambo	1.0300	-78.5800
10	ANSP	2315	Chocó	*X. m. littoralis*	Ecuador	Esmeraldas	20 km NNW Alto Tambo	1.0300	-78.5800
11	LSUMZ	11948	Chocó	*X. m. littoralis*	Ecuador	Esmeraldas	El Placer	0.8667	-78.5500
12	LSUMZ	28753	Chocó	*X. m. ridgwayi*	Panama	Colón	Road S-9 W off Gatun-Escobal Road (S-10), ca. 6 Kilometers SW Gatun	9.2800	-79.7100
13	LSUMZ	2209	Chocó	*X. m. littoralis*	Panama	Darién	Cana on E slope Cerro Pirré	7.7560	-77.6840
14	LSUMZ	26932	Chocó	*X. m. ridgwayi*	Panama	Panamá	Old Gamboa Road, 5 km NW Paraiso	9.0583	-79.6508
15	UWBM	jmd270	Chocó	*X. m. ridgwayi*	Panama	Panamá	Chagres National Park (old boyscout camp)	9.2500	-79.5830
16	UWBM	gms1842	Chocó	*X. m. ridgwayi*	Panama	Panamá	20 km ESE Canita, Lago Bayano	9.1532	-78.6929
17	USNM	5132	Guiana	*X. m. ruficaudus*	Guyana	Essequibo	Waruma River, E bank, ca. 15 river km S Kako River	5.5000	-60.7833
18	USNM	10887	Guiana	*X. m. ruficaudus*	Guyana	Northwest	North Side Acari Mountains	1.3833	-58.9333
19	USNM	9333	Guiana	*X. m. ruficaudus*	Guyana	Northwest	Baramita	7.3667	-60.4833
20	KUMNH	3879	Guiana	*X. m. ruficaudus*	Guyana	Cuyuni-Mazaruni	N slope Mount Roraima	5.2167	-60.7500
21	LSUMZ	45809	Guiana	*X. m. ruficaudus*	Suriname	Sipaliwini	Lely Gegberte	4.2744	-54.7391
22	AMNH	12699	Guiana	*X. m. ruficaudus*	Venezuela	Amazonas	Rio Baria, Cerro de la Neblina base camp	0.8342	-66.1667
23	AMNH	8845	Guiana	*X. m. ruficaudus*	Venezuela	Amazonas	Mrakapiwie	1.8954	-65.0456
24	AMNH	11942	Guiana	*X. m. ruficaudus*	Venezuela	Bolivar	40 km E Tumaremo on road to Bochinche	7.3833	-61.2167
25	FMNH	456908	Napo	*X. m. obsoletus*	Brazil	Amazonas	Japurá, Rio Mapari	-2.0497	-67.2631
26	FMNH	456909	Napo	*X. m. obsoletus*	Brazil	Amazonas	Japurá, Rio Mapari	-2.0497	-67.2631

APPENDIX B CONT.

27	MPEG	JAP 231	Napo	*X. m. obsoletus*	Brazil	Amazonas	Japurá, Rio Mapari	-2.0421	-67.2879
28	MPEG	JAP 299	Napo	*X. m. obsoletus*	Brazil	Amazonas	Japurá, Rio Mapari	-2.0421	-67.2879
29	ANSP	1484	Napo	*X. m. obsoletus*	Ecuador	Morona-Santiago	Santiago	-3.4000	-78.5500
30	LSUMZ	4244	Napo	*X. m. obsoletus*	Peru	Loreto	Lower Rio Napo, E bank Rio Yanayacu, ca. 90 km N Iquitos	-2.8200	-73.2738
31	LSUMZ	6862	Napo	*X. m. obsoletus*	Peru	Loreto	5 km N Amazon River, 85 km NE Iquitos	-3.4167	-72.5833
32	LSUMZ	7127	Napo	*X. m. obsoletus*	Peru	Loreto	5 km N Amazon River, 85 km NE Iquitos	-3.4167	-72.5833
33	LSUMZ	9026	Inambari	*X. m. obsoletus*	Bolivia	Pando	Nicolás Suarez, 12 km by road S Cobija, 8 km W on road to Mucden	-11.4703	-68.7786
34	MPEG	ESEC 225	Inambari	*X. m. obsoletus*	Brazil	Acre	ESEC Rio Acre, ca. 78 km W Assis, Brasil	-11.0568	-70.2713
35	MPEG	UFAC 1858	Inambari	*X. m. obsoletus*	Brazil	Acre	Feijó, Rio Envira, Novo Porto, Foz do Ig. Paraná do Ouro	-8.4599	-70.5564
36	MPEG	UFAC 815	Inambari	*X. m. obsoletus*	Brazil	Acre	Rio Branco, Transacreana (AC-090) km 70, Ramal Jarinal km 11	-9.9006	-68.4756
37	MPEG	UFAC 879	Inambari	*X. m. obsoletus*	Brazil	Acre	Rio Branco, Transacreana (AC-090) km 70, Ramal Jarinal km 11	-9.9006	-68.4756
38	MPEG	PUC 131	Inambari	*X. m. obsoletus*	Brazil	Amazonas	Tefé, Base Petrobras/Urucu, Papagaio	-4.8500	-65.0667
39	KUMNH	18530	Inambari	*X. m. obsoletus*	Peru	Cusco	ca. Alto Manguriari	-12.5655	-73.0878
40	FMNH	433364	Inambari	*X. m. obsoletus*	Peru	Cusco	Consuelo, 15.9 km SW Pilcopata	-13.0167	-71.4833
41	FMNH	391109	Rondónia	*X. m. obsoletus*	Bolivia	Beni	Hacienda Los Angeles, 10 km E Riberalta	-11.0092	-65.9952
42	LSUMZ	14752	Rondónia	*X. m. obsoletus*	Bolivia	Santa Cruz	Serrania de Huanchaca, 25km SE Calorata Arco Iris	-14.4867	-60.6753
43	LSUMZ	15114	Rondónia	*X. m. obsoletus*	Bolivia	Santa Cruz	Velasco, 13 km SW Piso Firme	-13.7700	-61.9500
44	LSUMZ	18175	Rondónia	*X. m. obsoletus*	Bolivia	Santa Cruz	Velasco, Parque Nacional Noel Kempff Mercado 86 km ESE Florida	-14.8333	-60.4167
45	LSUMZ	18534	Rondónia	*X. m. obsoletus*	Bolivia	Santa Cruz	Velasco, Parque Nacional Noel Kempff Mercado 60 km ESE of Florida	-14.8400	-60.7300
46	MPEG	FPR 040	Rondónia	*X. m. obsoletus*	Brazil	Amazonas	Maués, Flona do Pau Rosa, Comunidade Fortaleza	-3.9461	-58.4561
47	MPEG	FPR 103	Rondónia	*X. m. obsoletus*	Brazil	Amazonas	Maués, Flona do Pau Rosa, Comunidade Sta. Teresa	-3.4000	-57.7000
48	MPEG	MPDS 650	Rondónia	*X. m. obsoletus*	Brazil	Amazonas	Município de Humaitá, T. Indígena Parintintin, Aldeia Pupunha, Castanhal	-7.4667	-62.8167
49	MPEG	DED 323	Tapajós	*X. m. genibarbis*	Brazil	Mato Grosso	Município Nova Bandeirante, right bank Rio Juruena, Fazenda Vale Verde	-10.2519	-58.2850
50	FMNH	392023	Tapajós	*X. m. genibarbis*	Brazil	Mato Grosso	Municipio Alta Floresta, upper Rio Teles Pires-Rio Cristalino	-9.9040	-55.8810
51	MPEG	BR163-070	Tapajós	*X. m. genibarbis*	Brazil	Pará	Altamira, 30 km SW Castelo dos Sonhos, Fazenda Jamanxin	-8.3894	-55.3702
52	MPEG	BR163-181	Tapajós	*X. m. genibarbis*	Brazil	Pará	Itaituba, 7 km NW Moraes de Almeida	-6.2021	-55.6882
53	MPEG	FLJA 029	Tapajós	*X. m. genibarbis*	Brazil	Pará	Novo Progresso, margem esquerda Rio Jamanxim	-4.7000	-56.4500
54	MPEG	PIME 017	Tapajós	*X. m. genibarbis*	Brazil	Pará	Belterra, Flona do Tapajós, Br 163 km 117	-2.6333	-54.9500
55	MPEG	PIME 131	Tapajós	*X. m. genibarbis*	Brazil	Pará	Placas, Assentamento Comunidade Fortaleza	-3.4729	-54.5655

APPENDIX B CONT.

56	MPEG	WM344	Tapajós	*X. m. genibarbis*	Brazil	Pará	Belterra, Flona do Tapajós, Santarém/Cuiabá, BR 163 Km 117	-3.3561	-54.9492
57	FMNH	391347	Xingu	*X. m. genibarbis*	Brazil	Pará	Serra dos Carajas	-6.0783	-50.2468
58	FMNH	391348	Xingu	*X. m. genibarbis*	Brazil	Pará	Serra dos Carajas	-6.0783	-50.2468
59	FMNH	456904	Xingu	*X. m. genibarbis*	Brazil	Pará	Portel, FLONA do Caxiuanã, Plot PPBIO	-1.9500	-51.6000
60	FMNH	456905	Xingu	*X. m. genibarbis*	Brazil	Pará	Portel, FLONA do Caxiuanã, Plot PPBIO	-1.9500	-51.6000
61	FMNH	456906	Xingu	*X. m. genibarbis*	Brazil	Pará	Portel, FLONA do Caxiuanã, Plot PPBIO	-1.9500	-51.6000
62	MPEG	FTA 023	Xingu	*X. m. genibarbis*	Brazil	Pará	Carajás, FLONA Tapirapé-Aquiri	-2.9500	-51.8667
63	MPEG	MOP 048	Xingu	*X. m. genibarbis*	Brazil	Pará	Ourilandia do Norte, Serra do Puma	-6.7490	-51.0814
64	MPEG	PPBIO 151	Xingu	*X. m. genibarbis*	Brazil	Pará	Portel, FLONA do Caxiuanã, Plot PPBIO	-1.9500	-51.6000
65	MZUSP	1667	Atlantic	*X. m. minutus*	Brazil	São Paulo	Fazenda Barreiro Rico, São Paulo	-23.7114	-47.4188
66	MZUSP	685	Atlantic	*X. m. minutus*	Brazil	São Paulo	Piedade	-23.7114	-47.4188
67	MZUSP	689	Atlantic	*X. m. minutus*	Brazil	São Paulo	Piedade	-23.7114	-47.4188
68	KUMNH	255	Atlantic	*X. m. minutus*	Paraguay	Caazapá	San Rafael National Park	-26.3796	-55.6456
69	KUMNH	293	Atlantic	*X. m. minutus*	Paraguay	Caazapá	San Rafael National Park	-26.3796	-55.6456
70	KUMNH	342	Atlantic	*X. m. minutus*	Paraguay	Caazapá	San Rafael National Park	-26.3796	-55.6456
71	KUMNH	373	Atlantic	*X. m. minutus*	Paraguay	Caazapá	San Rafael National Park	-26.3796	-55.6456
72	LSUMZ	25938	Atlantic	*X. m. minutus*	Paraguay	Caazapá	Cord. de Caaguazu, 7.5 km E San Carlos	-26.1000	-55.7667

APPENDIX B CONT. Options used in the UNEAK pipeline for data processing.

Plug-in	Option	Value	Description
UMergeTaxaTagCountPlugin	-m	200000000	Maximum tag number in the merged TagCount file. Default: 60000000
UmergeTaxaTagCountPlugin	-c	5	Minimum count of a tag must be present to be output. Default: 5
UmergeTaxaTagCountPlugin	-t		Merge identically named taxa or not. -t n = do not merge. Default: merge
UTagCountToTagPairPlugin	-e	0.03	Error tolerance rate in the network filter. Default: 0.03
UMapInfoToHapMapPlugin	-mnMAF	0.05	Minimum minor allele frequency. Default: 0.05
UMapInfoToHapMapPlugin	-mxMAF	0.5	Maximum minor allele frequency. Default: 0.5
UMapInfoToHapMapPlugin	-mnC	0	Minimum call rate (proportion of taxa covered by at least one tag)
UMapInfoToHapMapPlugin	-mxC	1	Maximum call rate. Default: 1 (proportion of taxa covered by at least one tag)

APPENDIX B CONT. Processing statistics from the UNEAK pipeline.

	Mean	Median	Standard Deviation
Individual (Taxa) Depth	5.1253	5.0640	1.3068
Site Depth	4.9402	3.9251	4.4381
Individual (Taxa) Missingness	0.6776	0.6744	0.0528
Site Missingness	0.6776	0.8000	0.3003

APPENDIX B CONT. Results of aligning GBS loci to the Zebra Finch (*Taeniopygia guttata*) genome (number of loci with the best-scoring blastn hit falling on each *T. guttata* chromosome).

Zebra Finch Chromosome	Number of loci with highest-scoring blastn hit	Assembly Size (Mb) in Zebra Finch
1	162	118.550
1A	186	73.660
1B	3	1.080
2	248	156.410
3	250	112.620
4	141	69.780
4A	99	20.700
5	156	62.380
6	96	36.310
7	94	39.840
8	97	27.990
9	87	27.240
10	68	20.810
11	73	21.400
12	66	21.580
13	78	16.960
14	84	16.420
15	74	14.430
16	0	0.010
17	56	11.650
18	71	11.200
19	58	11.590
20	66	15.650
21	40	5.980
22	15	3.370
23	38	6.200
24	38	8.020
25	9	1.280
26	36	4.910
27	28	4.620
28	35	4.960
LGE22	0	0.883
LG2	0	0.110
LG5	0	0.016
Z	142	72.860
MT	0	0.017
Unknown	553	174.340

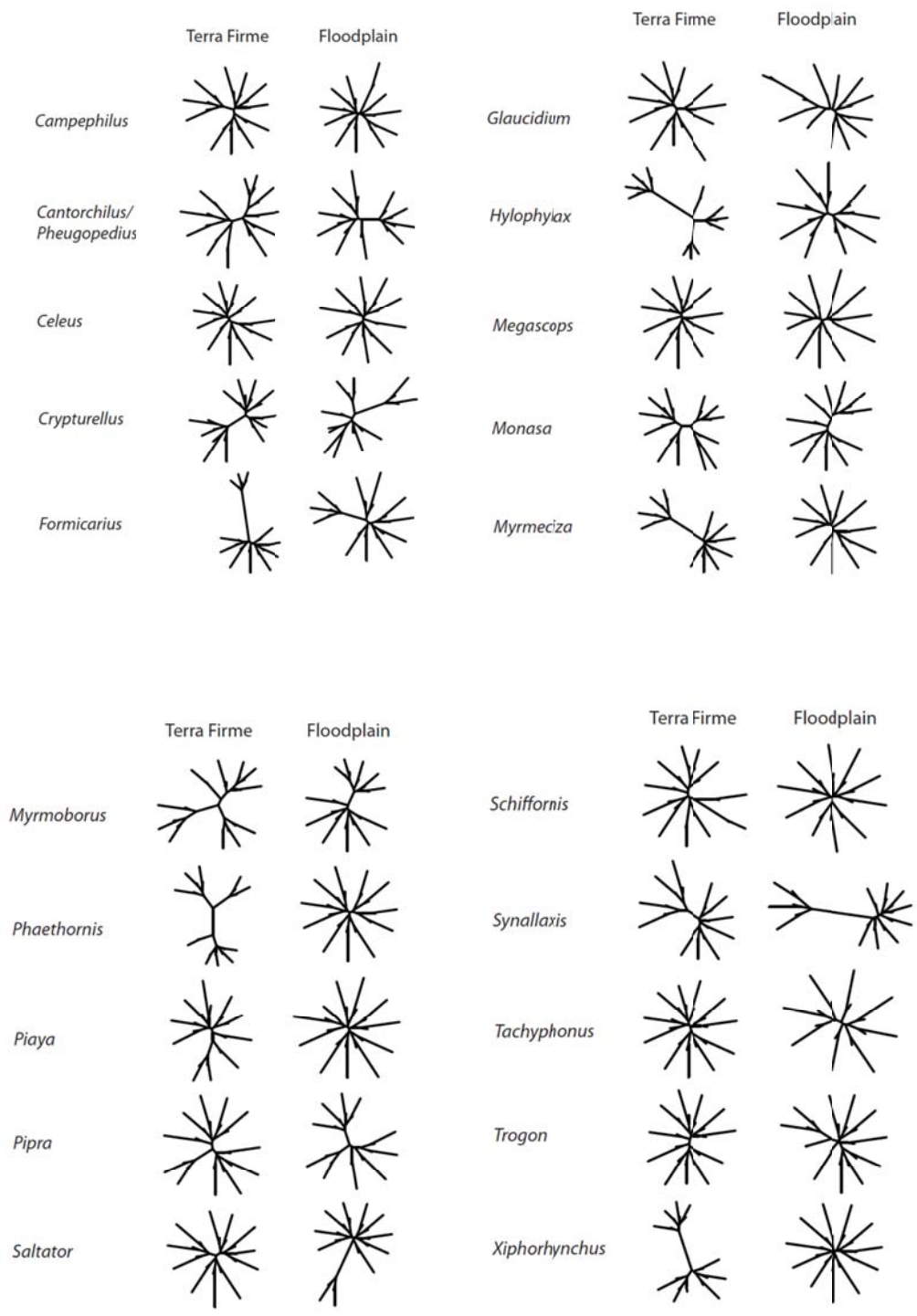
APPENDIX B CONT. Mantel and partial Mantel test results. A dash (-) separates the variables being examined, while a comma (,) precedes the variable being controlled for in partial Mantel tests.

Dataset	Test	r-statistic (95% CI)	p-value
Isolation by Distance			
All areas	Mantel (Geography - Fij)	-0.4964 (-0.5211, -0.4783)	0.0001*
All areas	Partial Mantel (Geography - Fij, Barriers)	-0.3133 (-0.3461, -0.2860)	0.0001*
Central America	Mantel (Geography - Fij)	-0.1225 (-0.4487, 0.1153)	0.3485
Chocó	Mantel (Geography - Fij)	-0.3425 (-0.6126, -0.0575)	0.0605
Guiana	Mantel (Geography - Fij)	-0.3588 (-0.5673, -0.1277)	0.1769
Napo	Mantel (Geography - Fij)	-0.4069 (-0.4741, -0.3612)	0.0081*
Inambari	Mantel (Geography - Fij)	-0.2762 (-0.5604, 0.2833)	0.3738
Rondônia	Mantel (Geography - Fij)	-0.5859 (-0.8680, -0.3955)	0.0072*
Tapajós	Mantel (Geography - Fij)	-0.1646 (-0.4353, 0.0434)	0.4317
Xingu	Mantel (Geography - Fij)	-0.2824 (-0.4095, -0.02415)	0.1105
Atlantic Forest	Mantel (Geography - Fij)	-0.5816 (-0.8176, -0.3454)	0.0032*
Isolation by Barriers			
All barriers	Partial Mantel (Barrier - Fij, Geography)	-0.6467 (-0.6762, -0.6123)	0.0001*
Isthmus of Panama	Partial Mantel (Barrier - Fij, Geography)	-0.7158 (-0.8085, -0.6461)	0.0001*
Andes Mountains	Partial Mantel (Barrier - Fij, Geography)	-0.7373 (-0.7978, -0.6203)	0.0001*
Rio Negro	Partial Mantel (Barrier - Fij, Geography)	-0.7969 (-0.8432, -0.7362)	0.0001*
Rio Solimões	Partial Mantel (Barrier - Fij, Geography)	-0.5187 (-0.8303, -0.3586)	0.0001*
Rio Madeira	Partial Mantel (Barrier - Fij, Geography)	-0.4689 (-0.6611, -0.3568)	0.0015*
Rio Tapajós	Partial Mantel (Barrier - Fij, Geography)	-0.8435 (-0.9236, -0.7997)	0.0004*
Rio Xingu	Partial Mantel (Barrier - Fij, Geography)	-0.2756 (-0.4101, -0.1796)	0.0074*
Cerrado Belt	Partial Mantel (Barrier - Fij, Geography)	-0.5313 (-0.7121, -0.4212)	0.0002*
Population Validation			
STRUCTURE/Structurama populations	Partial Mantel (Populations - Fij, Geography)	-0.7611 (-0.7937, -0.7282)	0.0001*
fineSTRUCTURE populations	Partial Mantel (Populations - Fij, Geography)	-0.6709 (-0.7167, -0.6293)	0.0001*

* $P<0.01$

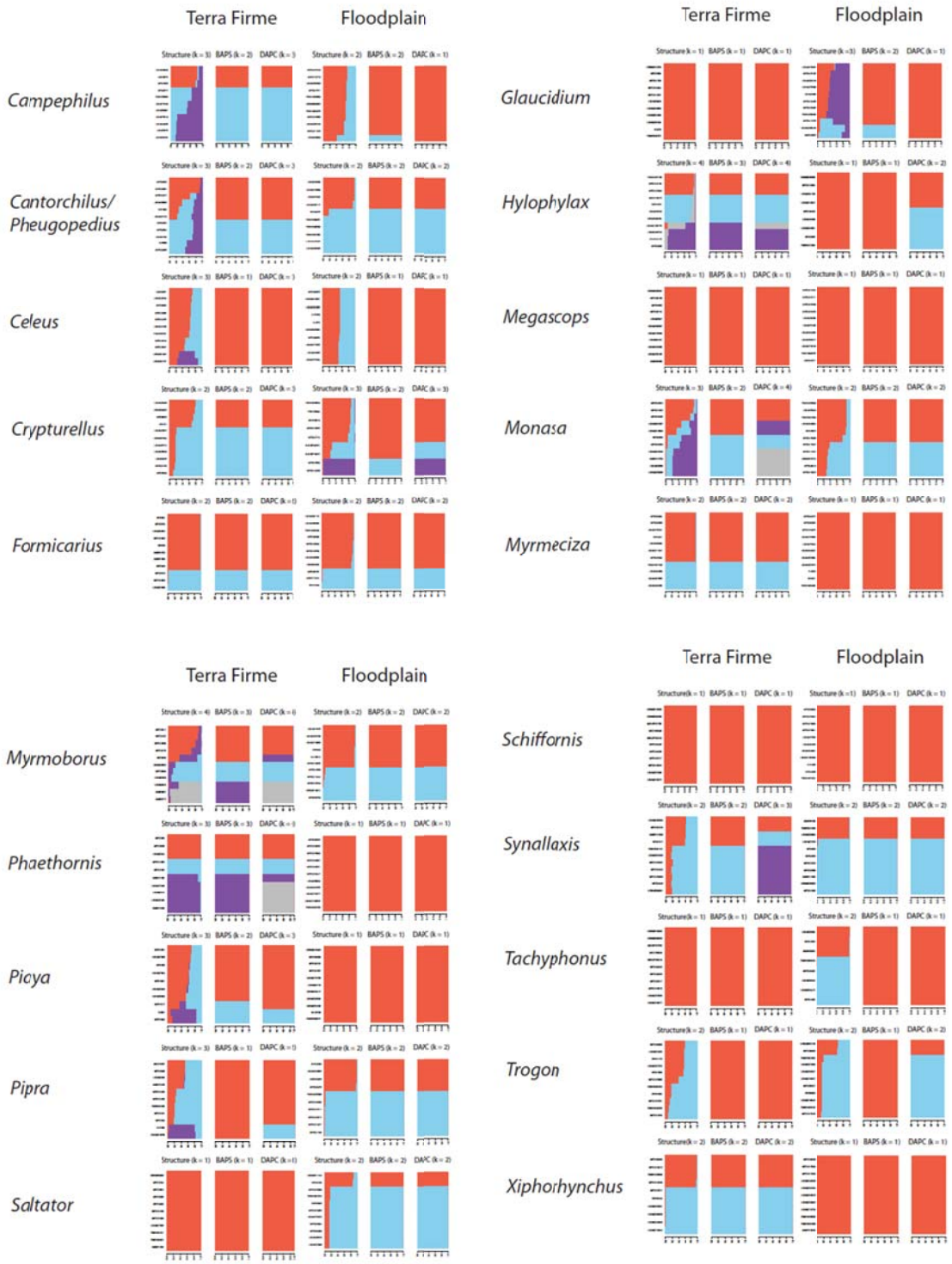
APPENDIX B CONT. Results from STRUCTURE runs.

K	Pr(X|K)	Pr(K)
1	-167422.6	-166580.1
2	-128202.2	-126761.2
3	-100979.9	-99679.6
4	-77017.2	-75504
5	-65265.1	-63458.3
6	-77045.9	-75514.4
7	-77065	-75520
8	-65323.2	-63480.1
9	-65352	-63487.1
10	-65366.5	-63494
11	-77116.3	-75539.6
12	-65402.3	-63506.8
13	-65423.1	-63513.7
14	-65440.2	-63520.6
15	-65452.1	-63526.5



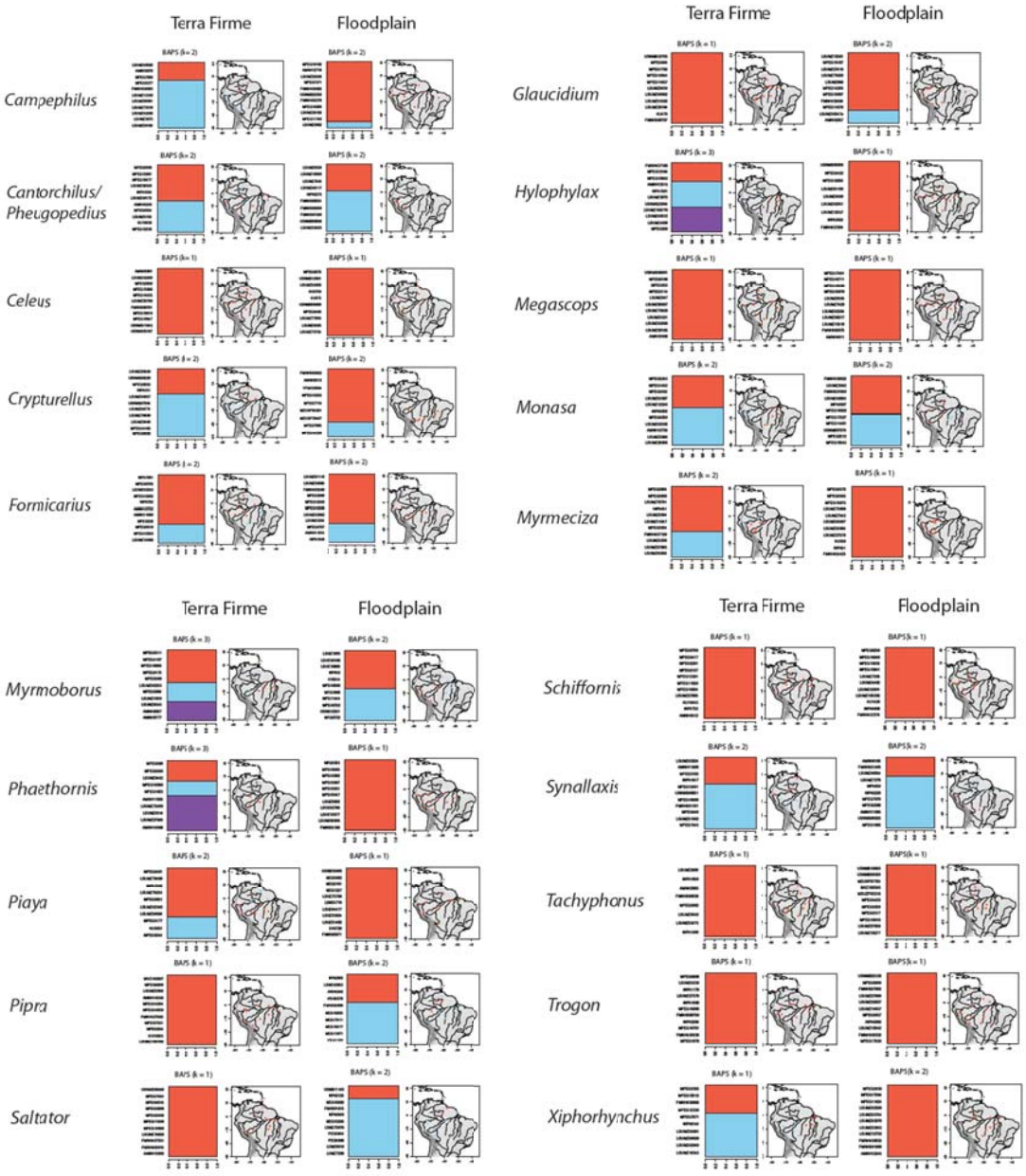
Bayesian phylogenetic trees of concatenated SNPs for ingroups in each species.

APPENDIX C CONT.



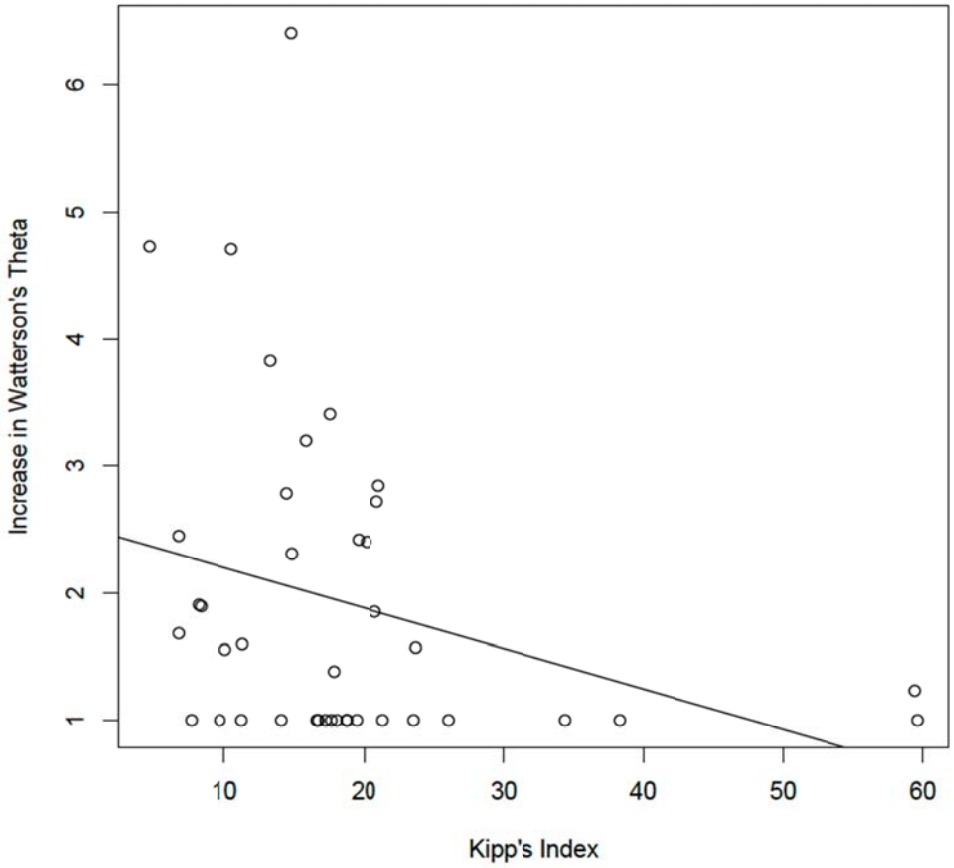
Population clustering results from STRUCTURE, BAPS, and DAPC for ingroup in each species.

APPENDIX C CONT.



BAPS population clustering results and maps of BAPS populations for ingroup of each species.

APPENDIX C CONT.



Relationship between Kipp's Index and Watterson's theta across study species.

APPENDIX D: SUPPLEMENTARY MATERIAL TO CHAPTER 5

Sampling and Sequencing

I selected 177 species for examination. I sampled only mainland New World taxa to help control for the area available to each species for accruing allopatric divergence. In addition, New World bird species are better represented in existing genetic resources collections and genetic datasets (Reddy 2014). Species that had distributions extending into the Old World were included, but samples from Old World populations were not examined. Differences among geographic regions in the number of species having undergone recent taxonomic revision and in the application of different taxonomic criteria may result in biased results, thus I used a standardized set of criteria to circumscribe species for the purposes of this study (but see below for examination of an alternative taxonomy). Species were defined as all non-sympatric monophyletic populations for which I had sampling, regardless of their current treatment by taxonomic authorities. They therefore represent "lumped" species or superspecies.

For each species, I selected at least 8 samples (mean = 111) distributed widely geographically. I extracted whole DNA from tissue samples associated with voucher specimens using Qiagen DNeasy extraction kits (Valencia, CA) following the manufacturer's protocol. I used polymerase chain reaction to amplify sequence from the mitochondrial genes NADH dehydrogenase 2 (ND2) using standard primers. I conducted Sanger sequencing on PCR amplicons and aligned and edited sequences using I also collected population-level mitochondrial datasets from Genbank

for existing studies of New World birds, again restricting sampling to datasets containing at least 8 samples (mean = 95).

Population Divergence Estimation

We estimated mitochondrial gene trees for each species using the Bayesian method implemented in BEAST v.1.7.5 (Drummond et al. 2012). All trees were time-calibrated using an uncorrelated relaxed substitution rate based on published avian mitochondrial rates of 0.0125 substitutions/site/My for ND2 and ATPASE6, 7, & 8 (Smith and Klicka 2010) and 0.0105s s/s/My for cyt *b* (Weir and Schluter 2008). For the gene COI I used the same rate as cyt *b* because the loci mutate at similar rates (Smith and Klicka 2010). For the clock rate parameter I specified a lognormal normal distribution on the prior with the mean set to the above-mentioned mutation rates and a standard deviation of 0.1. I used a coalescent-constant-size tree prior and the best-fit nucleotide substitution model as determined in MEGA6 (Tamura et al. 2013). I ran each analysis for 50 million generations sampling every 2,500 generations, performed multiple independent runs for validation, and assessed MCMC convergence and determined burn-in by examining ESS values and likelihood plots in Tracer v.1.5 (Rambaut and Drummond 2007). For some datasets that did not achieve high ESS values after 50 million generations, I included additional generations until the results were stable. I included taxa deemed to be sister to study species based on prior phylogenetic work and I extracted stem and crown age estimates for each species.

We quantified phylogeographic structure using a Bayesian implementation of the General Mixed Yule Coalescent model (bGMYC; Reid and Carstens 2012). bGMYC determines the number of genetic species by estimating the number of clusters within which splits in the gene tree fit a coalescent model rather than a model of interspecific diversification (Yule model). Populations of most birds are structured to the extent that many taxonomic species contain multiple geographically separated bGMYC clusters or genetic species. I used the maximum clade credibility (MCC) tree from BEAST for each bGMYC run. I ran the program for 250,000 generations using the single.phy function and discarded the first 15,000 generations as burn-in. I ran each analysis multiple times for validation, and assessed MCMC diagnostics by examining likelihood plots in Tracer. bGMYC provides a posterior probability that two sequences belong to the same species which can be used, along with a probability threshold, to determine the number of clusters. For the primary analysis I used a posterior probability threshold of 0.8 for clustering (but see exploration of this setting below).

We determined the rate of bGMYC cluster formation, hereafter the phylogeographic splitting rate, as an index of the rate at which geographic variation accrues in each species. I calculated rates using crown age, the time before present of the first intra-specific divergence event. I calculated rates of bGMYC cluster formation under a pure-birth model using formula (6) from Magallón and Sanderson (2001) as implemented in the R package laser (Rabosky and Schliep 2013). All rates were calculated using a starting diversity of one despite the use of crown age. Crown age in my study corresponds to the first divergence between mitochondrial haplotypes rather

than the first divergence between bGMYC clusters, and thus represents a time point when only one bGMYC cluster was present. See exploration of alternative approaches for rate estimation below.

Speciation Rate Estimation

We used time-calibrated MCC trees from a prior phylogeny of all birds (Jetz et al. 2012) for estimation of speciation rates. Jetz et al. constructed trees by estimating subtrees from genetic data for smaller clades, then placing them on one of two backbone phylogenies (Hackett et al. 2008, Ericson 2012). Jetz et al. placed species lacking genetic data using taxonomic constraints, but I removed these (leaving 6,670 species) for my analyses to eliminate potential artifacts due to incorrect placement and because the BAMM model I used to analyze speciation rates incorporates an explicit analytical correction for incomplete taxon sampling.

We estimated speciation rates in the pruned MCC trees based on the Hackett et al. backbone using the model implemented in the program BAMM (Rabosky et al. 2013, Rabosky 2014). BAMM uses reversible-jump Markov Chain Monte Carlo to examine models differing in the number of time-varying diversification processes present across the phylogeny. Each process includes a time-varying speciation term and a time-invariant extinction rate. BAMM was run assuming 67% sampling across the avian tree to account for species without genetic data. Speciation rates for a given terminal branch on the tree

were extracted from the marginal distribution of rates, which is based on all processes sampled at that branch.

We ran BAMM for at least 350 million generations in each analysis, completing multiple runs with the same settings for validation. I sampled every 200,000 generations and discarded 10% of the sample as burn-in. I extracted the means of the marginal distributions of tip (present-day) speciation rates for all species for which I had phylogeographic data. Overall, there were 47 distinct macroevolutionary regimes represented in the full phylogeny of 6,670 species. Within the tree containing only the 177 study species, there were 23 regimes. However, some of these regimes occurred along the same branch, due to the excision of lineages that contributed to rate shifts but were not represented in my dataset. Thus, the sampled tips were subtended by 21 regimes, including the "background" regime beginning at the base of the tree.

Comparative Analyses

We examined correlations between phylogeography and speciation rate using a semi-parametric trait-dependent diversification test that detects effects based on replicated associations between trait values and diversification rates (Rabosky and Huang in press). This test computes the correlation between character states at the tips of the tree and their corresponding diversification rates and assesses significance by permuting speciation rates among regimes. Parametric uncertainty in diversification rates is accommodated by conducting tests across the posterior distribution of rates inferred using

BAMM. The permutation test is used to control for the covariance among species from the same macroevolutionary rate regime, thereby explicitly incorporating covariance among replicates with shared history and macroevolutionary dynamics. I tested the Type I error rate of this approach by simulating trait evolution on my tree under a Brownian motion model (500 replicates) and conducting the trait-dependent diversification test on the simulated data. I compared the Type I error rate using the trait-dependent diversification test to that from traditional Spearman's rank-correlation without accounting for the covariance structure of my data. The Type 1 error rate for the lumped taxonomy was 0.038 (relative to 0.522 from the Spearman's rank correlation) and for the split taxonomy (see below) was 0.046 (relative to 0.566 from the Spearman's rank correlation).

Use of Mitochondrial Versus Genomic Data

Currently, the only population-level genetic datasets encompassing sufficient species for large comparative analyses comprise mitochondrial DNA (mtDNA) sequences, and I were therefore limited to examining mtDNA for this study. my results assume that mitochondrial gene trees provide an accurate indication of the number and age of divergent populations within a species. Although mitochondrial DNA has long been used as a marker for population-level studies, mitochondrial gene trees may not capture the population history of a species in cases when gene flow occurs between populations or in cases of incomplete lineage sorting (Brito and Edwards 2009). The

number of species and mitochondrial gene trees examined in this study should overcome much of the random stochasticity in individual gene genealogies. Deterministic biases could be problematic for my analyses, however, if mitochondrial gene trees are biased with respect to the speciation rate of their ancestral lineage. If the number of inferred bGMYC clusters is upward biased in shallow gene trees, which are typical of clades with rapid diversification, this bias could lead to the observed relationship between rate of population divergence and diversification rate. Simulation studies, however, reveal that, although the accuracy of bGMYC decreases when gene trees are simulated within shallower species trees, there is no deterministic bias in the number of clusters inferred under different tree depths (Reid and Carstens 2012). Thus, bGMYC estimates of cluster number from mitochondrial gene trees are unlikely to be biased in a deterministic manner relative to diversification rate.

To further examine the level to which bGMYC clusters inferred from mitochondrial data accurately reflect population structure, I examined genomic datasets of single nucleotide polymorphisms from the nuclear genome of *Xenops minutus*. The methods used to obtain these data are presented in Harvey and Brumfield (2015). Briefly, I sent samples from 72 individuals of *X. minutus* to the Cornell Institute of Genomic Diversity, where they were prepared using a RAD-Seq protocol (Elshire et al. 2011) and sequenced using an Illumina HiSeq 2000 massively parallel sequencer. I assembled reads using the UNEAK pipeline (Bradbury et al. 2007) and called single-nucleotide polymorphisms (SNPs) using the method of Lynch (2009) as implemented in custom perl scripts (White et al. 2013, https://github.com/mgharvey/GBS_process_Tom_White/v1).

We clustered individuals into populations using STRUCTURE 2.3.4 (Pritchard et al. 2000) examining all values of K (number of clusters) between 1 and 15. I examined the correspondence between number of STRUCTURE clusters from the genomic SNPs and number of bGMYC clusters from mitochondrial data from the same samples. Unfortunately, no standardized substitution rates are available for the genomic markers examined, so I were unable to assess the age of divergence events between populations within these species.

STRUCTURE analysis of 3,379 SNPs revealed a best-fit value of K of 5, based on the values of P(X|K). This is fewer clusters than the bGMYC results, which were 8 clusters based on a posterior probability threshold for clustering of 0.9, or 9 clusters based on posterior probability thresholds of 0.8 or 0.7 (Appendix B). All breaks present in the STRUCTURE analysis, however, were also present in the bGMYC analysis. The bGMYC analyses, however, resulted in the subdivision of some STRUCTURE clusters. This result may reflect greater sensitivity of the bGMYC program, or the faster time to reciprocal monophyly of mitochondrial sequences (Palumbi et al. 2001). These results demonstrate that mitochondrial gene trees capture similar patterns of structure to genomic datasets, and that cluster membership is consistent between both classes of markers. Overall, these analyses support prior studies that found that mitochondrial datasets, despite their limitations, are adequate for comparative studies of divergence times and population structure across lineages (Smith et al. 2014).

Examination of Alternative Taxonomy

The estimation of population divergence requires that I circumscribe the populations included in the analysis for a given species. A more inclusive definition of species will generally result in more population clusters than a less inclusive definition. If species are defined differently amongst clades that differ in their speciation rates, the taxonomy used to circumscribe species may bias my results. I first alleviated this issue by focusing on rates of divergence rather than raw number of bGMYC clusters. I expect divergence rates to be similar in a species regardless of the taxonomic treatment, because a more inclusive treatment will generally result in older stem and crown ages for that species in addition to more bGMYC clusters.

We also investigated the impact of taxonomic treatment on results by examining two different taxonomies. The main taxonomy represented all allopatric populations that formed a monophyletic group, regardless of their current treatment by taxonomic authorities. A second taxonomy corresponded to the current taxonomy of the American Ornithologist's Union (AOU) North American (AOU 1998, Chesser et al. 2013) and South American (Remsen et al. 2014) checklist committees. In situations where the North and South American committees differed in their treatment, I reverted to the North American committee's treatment. The AOU taxonomy is more subdivided or "split" (208 species) than the primary taxonomy (177 species), so examination of both provides an

index of the impact of the level of taxonomic splitting on results. The split taxonomy resulted in a somewhat weaker but still significant relationship between population divergence and speciation rate (see Main Text).

Examination of Alternative Phylogenetic Tree

The phylogenetic tree used for primary analyses represents the first available phylogenetic hypothesis for all birds. It is possible, however, that my analyses might be biased by inaccuracies in this estimate of avian diversification history. To explore this possibility, I also estimated speciation rates using a second phylogenetic hypothesis for birds (Burleigh et al. in prep.). This phylogeny differs from that used in the main analysis in that it did not use taxonomic constraints to place taxa or involve a staged approach to tree estimation. 178 study species were present in this phylogeny. I found similar support for the association between population differentiation and speciation rate from this tree as in the primary analysis ($r = 0.247$, $p = 0.008$).

Examination of Sampling Adequacy

Although sampling was extensive for most species (mean = 100 samples), inadequate sampling could impact results if populations were missed in my datasets. If the number of populations missed corresponded to the speciation rate of a species, the

missing data could bias my results. I evaluated how robust my result was to missing data by randomly pruning 20% and 40% of the tips of the mitochondrial gene trees estimate from the full dataset, re-estimating the number of bGMYC clusters and rates of population divergence, and conducting trait-dependent diversification tests. Results were still significant at 20%, and weakly at 40% of the full dataset (see Main Text).

Examination of Alternative Posterior Probability Thresholds for Clustering in bGMYC

The Bayesian GMYC method used requires the researcher to set an arbitrary posterior probability threshold to determine the level of support required to cluster individuals into a single population. Higher thresholds generally result in fewer clusters than lower thresholds. To examine the impact of this threshold on results, I examined three different threshold values (0.9, 0.8, and 0.7). All three thresholds resulted in similar results (see Main Text).

Examination of divergence rate estimates using stem age. Divergence rates can be estimated either using the crown age of populations (the earliest divergence among populations within the species being examined) or the stem age of the populations (the time since the species examined diverged from it's nearest relative). Stem ages are always older than crown ages in gene trees of monophyletic species, thus divergence rates based on stem ages average slower than those based on crown ages. Crown age is generally superior to stem age for rate estimation because it is positively correlated with diversity

(Stadler et al. 2014), increasing the comparability of rate estimates across species and taxonomic treatments. I therefore focused on rates estimated using crown age for primary analyses, but I also examined rates from stem ages. Correlations between speciation rate and divergence rates based on stem ages were even stronger than those from divergence rates based on crown ages (see Main Text). This may be due to the fact that rates based on stem age incorporate some signal of extinction and persistence as well as the signal of divergence rate. Longer stem branches may be evidence of higher extinction (see below) and result in lower divergence rates estimating using stem age.

Examination of Birth-Death Models of Population Divergence

Because I modeled divergence at shallow time scales, I might assume that extinction is infrequent and pure-birth (Yule) models provide reasonable estimates of divergence rate. I tested this assumption by estimating divergence rate using birth-death models with moderate (epsilon = 0.45) and high (eps. = 0.9) rates of extinction, in addition to a pure-birth model. These resulted in lower rates of divergence than the pure-birth model, but correlations between divergence rate and speciation rate were similar to pure-birth using rate estimates incorporating both moderate and high extinction (see Main Text). I were unable to test models that jointly estimate divergence and extinction due to the small number of bGMYC clusters in many of the intraspecific datasets.

Examination of the Impact of Combining Datasets from Different Genes

The datasets examined represented different mitochondrial loci, including ND2 (n = 99), cyt *b* (n = 45), the mitochondrial control region (n = 17), ATPASE8 & 6 (n = 8), COI (n = 1), COII (n = 1), ND3 (n = 1), ND6 (n = 1), ATPASE8 & 7 (n = 1), or some combination thereof. To determine whether differences in gene histories among loci produced the observed correlations, I examined subsets of the dataset from ND2 and from cyt *b*, the two loci with the largest number of datasets. Trait-dependent diversification tests revealed correlations that were similar to the full dataset for ND2 (n = 99, r = 0.466, p = 0.002), although the relationship was non-significant amongst the relatively few species with cyt *b* datasets (n = 45).

Examination of the Impact of Including Single-Sample bGMYC Clusters

Some bGMYC clusters contained only a single sample. These were generally associated with long gene tree branches, which may not have been treated as a cluster if improved sampling resulted in additional branching events on that lineage. To examine this issue, I re-ran analyses after removing bGMYC clusters containing only a single sample. Results were similar, however, to those from the full dataset (r = 0.294, p = 0.004).

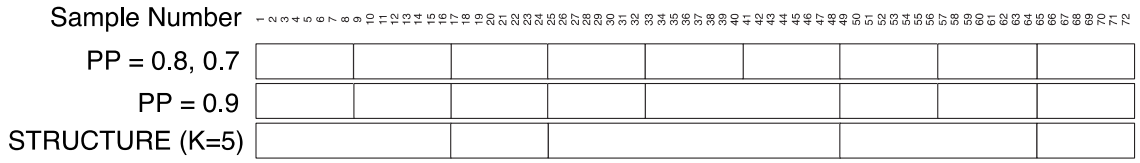
Population Persistence

The rate of persistence versus extinction of divergent populations may act as a control on speciation rate as well as their rate of divergence. my population histories generally contained too little information to jointly optimize population divergence and population extinction using a model. Instead, I used the relative length of the stem branch for each species as a rough index of extinction. The stem branch length reflects the time between the divergence between the study species and it's nearest relative, and the first divergence event with the study species. The length of this branch is expected to be positively correlated with the extinction rate of the species, all else being equal and assuming a constant rate of divergence (Nee et al. 1994, Dynesius and Jansson 2014). If persistence acts as a control on speciation rate, I expect a negative correlation between relative stem branch length and speciation rate. A trait-dependent diversification test on relative stem branch length revealed no significant correlation, however ($r = -0.167$, $p = 0.110$). The negative slope of this relationship is suggestive, but better information (e.g. fossil data) or more sophisticated metrics of extinction may be necessary to properly assess the impact of population persistence on macroevolutionary diversification.

Diversity-Dependence and Ecological Limits

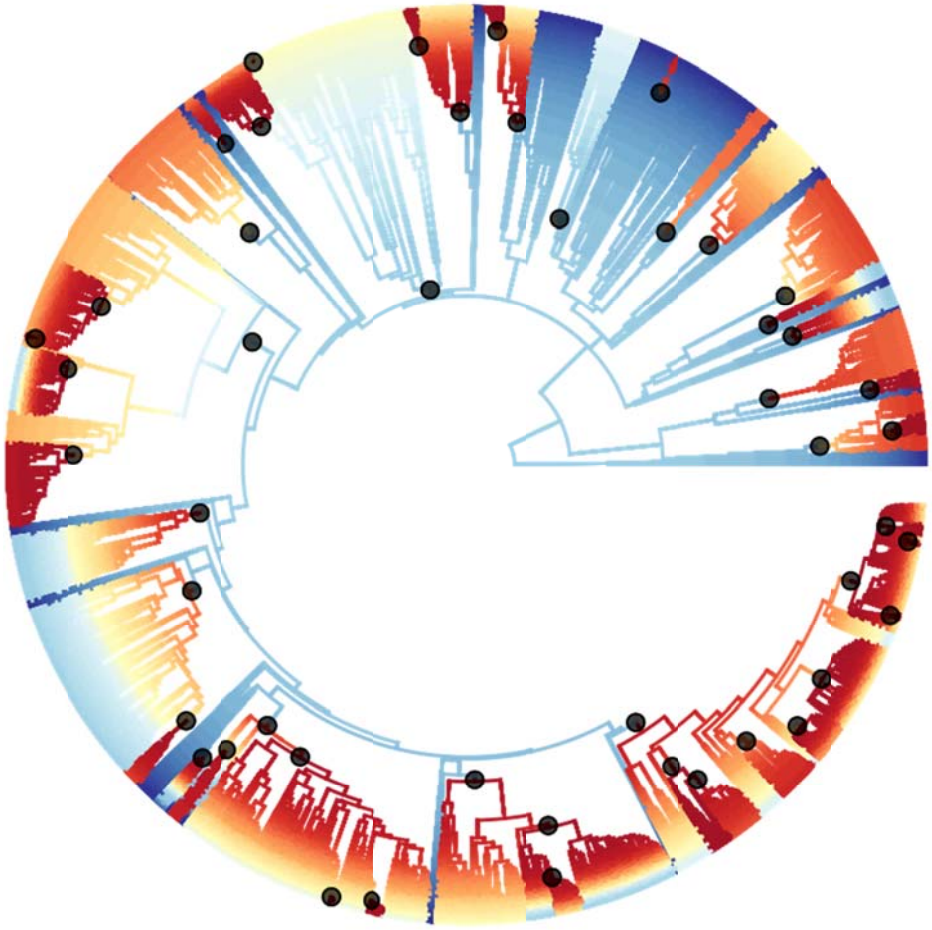
Ecological limits may impact speciation rates, for example by reducing rates in species that have saturated available niches (Valentine and Moores 1972, Rosenzweig 1975, Rabosky 2009), but may not have an impact on population divergence rate, most of which occurs between allopatric populations that are not in competition with each other. If this is the case, ecological limits, by limiting speciation rate only, may dampen the correlation between population divergence rate and speciation rate. To test this, I estimated the strength of diversity dependent relationships in each macroevolutionary regime recovered from the BAMM analyses. For each regime, I extracted the posterior mean speciation rate in the first time slice (1% of the duration of the regime) and the rate from the most recent time slice (also 1%). When the rate in the first time slice is faster than the rate in the most recent time slice this indicates a rate slowdown, which is typically ascribed to diversity dependent behavior (Rabosky 2009). I then examined the strength of the correlation between population divergence and speciation rates both in regimes with strong signals of diversity dependence (greater than a 2× slowdown in speciation rate), and those with weak or no signals of diversity dependence (less than or equal to a 2× slowdown). The threshold of 2× resulted in a relatively even split of number of regimes (9 and 12) and number of samples (83 and 94) between the diversity-dependent and non-diversity-dependent categories. As expected if ecological limits dampen the correlation between divergence and speciation, I found that there was no

correlation between population divergence and speciation rates in clades with signatures of diversity-dependence ($r = 0.027$, $p = 0.827$), but a relatively strong correlation in clades without diversity dependence ($r = 0.368$, $p = 0.026$). The latter is a stronger correlation than observed across all clades, suggesting that density dependence is reducing the correlation between divergence and speciation in my full dataset.

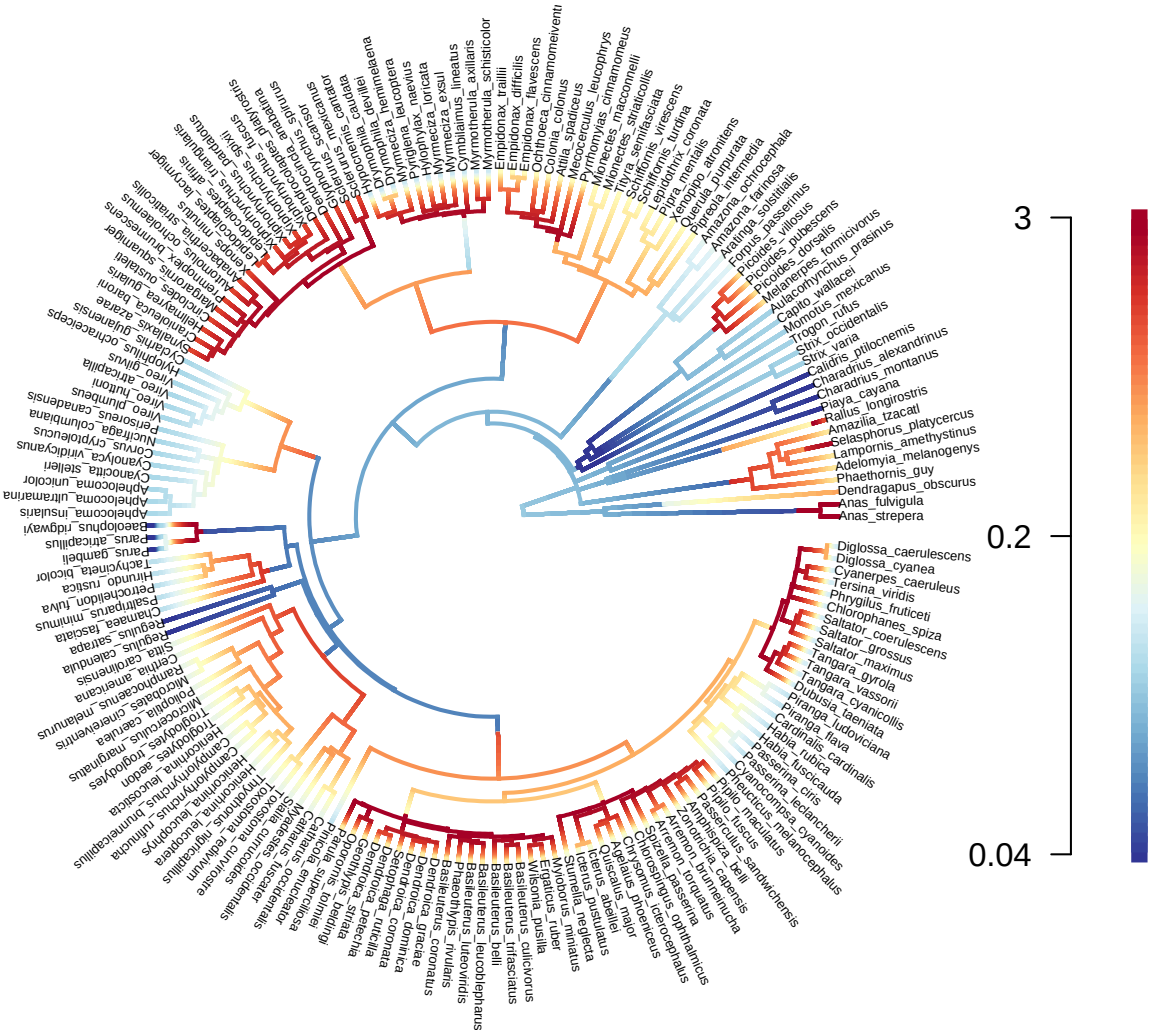


Population cluster assignment from bGMYC analysis of *Xenops minutus* mitochondrial gene tree with posterior probability threshold required for shared cluster membership set to 0.8/0.7 (top row) or 0.9 (middle row). The last row is population cluster assignment from STRUCTURE analysis of 3,379 genome-wide SNPs. The cluster membership is broadly concordant, although the mitochondrial gene trees combined with bGMYC result in finer subdivision than the STRUCTURE analysis of SNPs.

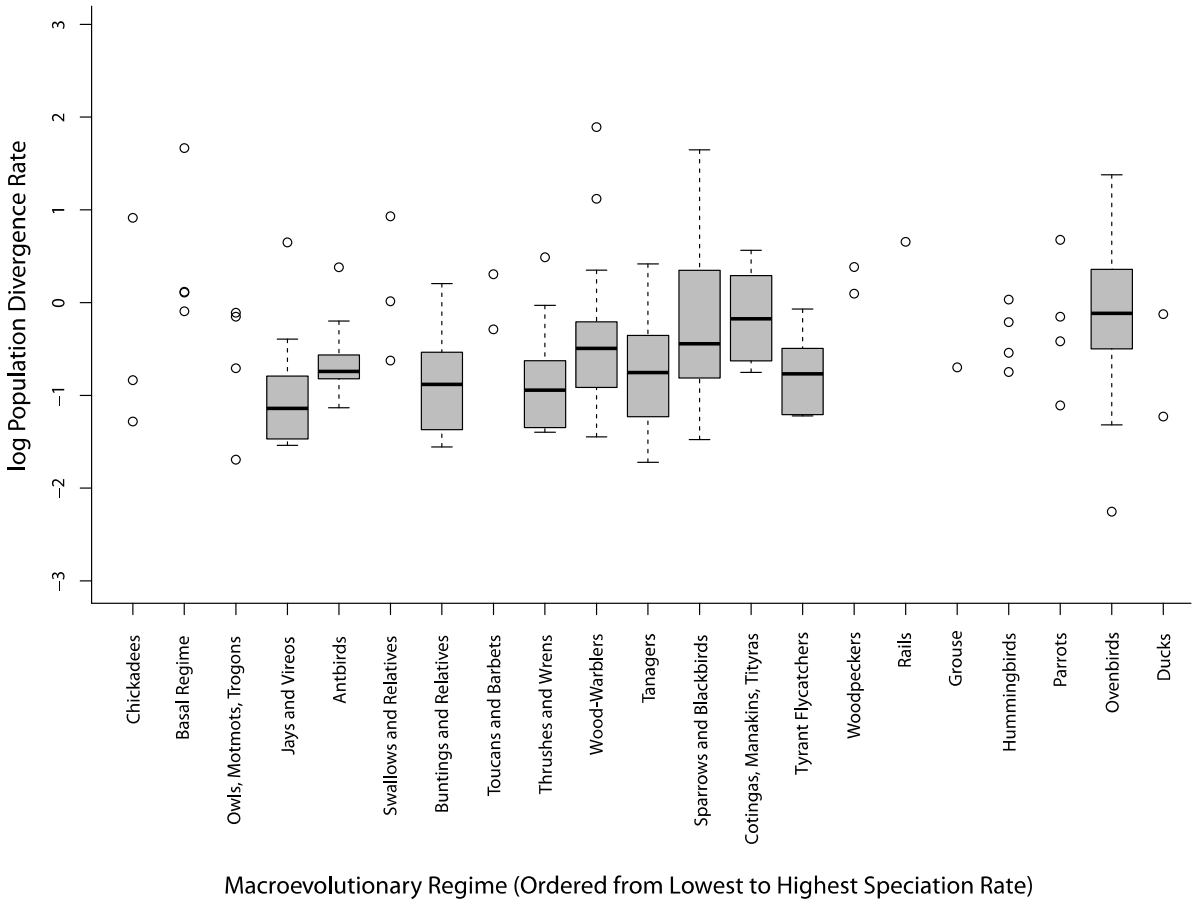
The distance between the oldest within-species haplotype divergence event and the stem age of a species (time since it split with its sister species), relative to the total time since the stem, serves as a proxy of the impact of extinction during the history of a lineage.



A phylorate plot showing speciation rates and regime shifts across the tree of all 6,670 birds with genetic data based on the Jetz et al. (2012) phylogeny.



The same phylorate plot from Figure 2 with individual tips labeled for reference. Colors indicate mean posterior speciation rates along branches.



Boxplot showing population divergence rates for each of the 21 terminal BAMM regimes containing study species. Rates from regimes containing six or fewer study species are plotted as individual points rather than boxes.

APPENDIX D CONT. The first column numbers the 177 study species, the second provides the species within each based on the American Ornithologist's Union (AOU) taxonomy, the third is the locus used for analysis, and the final column lists the source for each dataset (for previously published datasets).

Species	AOU Species	Locus	Author
1	*Adelomyia melanogenys*	CR-ATP8-ATP6	Chaves et al. 2007, Chaves and Smith 2011
2	*Agelaius phoeniceus* *Agelaius tricolor*	ND2	Barker et al. 2012
3	*Amazilia tzacatl*	ND2	Lelevier et al. 2011
4	*Amazona farinosa*	cyt	Wenner et al. 2012
5	*Amazona ochrocephala* *Amazona aestiva*	ND2	Eberhard et al. 2004, Ribas et al. 2007
6	*Anabacerthia striaticollis*	ND2	Cuervo et al. 2013
7	*Anas fulvigula*	CR	McCracken et al. 2001
8	*Anas strepera*	CR	Peters and Omland 2007
9	*Aphelocoma caerulescens* *Aphelocoma californica* *Aphelocoma insularis*	ND2	McCormack et al. 2011
10	*Aphelocoma ultramarina* *Aphelocoma wollweberi*	ND2	McCormack et al. 2008, McCormack et al. 2011
11	*Aphelocoma unicolor*	ND2	McCormack et al. 2011
12	*Aratinga solstitialis* *Aratinga jandaya* *Aratinga auricapillus*	CytB	Ribas and Miyaki 2004
13	*Arremon brunneinucha*	COII	Navarro-Sigüenza et al. 2008
14	*Arremon torquatus* *Arremon assimilis* *Arremon atricapillus* *Arremon basilicus* *Arremon costaricensis* *Arremon perijanus* *Arremon phaeopleurus* *Arremon phygas*	ND2	Cadena and Cuervo 2010; Cuervo et al. 2013
15	*Artemisiospiza belli* *Artemisiospiza nevadensis*	CytB	Cicero and Koo 2012
16	*Attila spadiceus*	CytB	Smith et al. 2014
17	*Aulacorhynchus prasinus*	ND2	Puebla-Olivares et al. 2008
18	*Automolus ochrolaemus*	CytB	Smith et al. 2014
19	*Baeolophus inornatus* *Baeolophus ridgwayi*	CytB	Cicero 2004
20	*Basileuterus belli*	ND2	Barber and Klicka 2010
21	*Basileuterus culicivorus*	CytB	Vilaca and Santos 2010

APPENDIX D CONT.

22	*Basileuterus tristriatus* *Basileuterus trifasciatus*	ND2	Gutierrez-Pinto et al. 2012
23	*Calidris ptilocnemis*	CytB	Pruett and Winker 2005
24	*Campylorhynchus brunneicapillus*	ND2	Zink et al. 2001
25	*Campylorhynchus rufinucha*	ND2	Vazquez-Miranda et al. 2009
26	*Cantorchilus nigricapillus*	ATP8-ATP6	Gonzalez et al. 2003
27	*Capito wallacei*	CytB	Seeholzer et al. 2011
28	*Cardellina pusilla*	CytB	Kimura et al. 2002
29	*Cardellina rubra* *Cardellina versicolor*	ND2	Barrera-Guzman et al. 2012
30	*Cardinalis cardinalis*	ND2	Smith et al. 2011
31	*Catharus fuscater*	ND2	Cuervo et al. 2013
32	*Catharus occidentalis*	ND2	Klicka et al.
33	*Ceratopipra chloromeros* *Ceratopipra mentalis* *Ceratopipra erythrocephala* *Ceratopipra rubrocapilla*	ND2	Harvey et al.
34	*Certhia americana*	ND2	Manthey et al. 2011
35	*Chamaea fasciata*	CytB	Burns and Barhoum 2006
36	*Charadrius montanus*	CR	Oyler-McCance et al. 2005, Funk et al. 2007
37	*Charadrius nivosus*	CR	Funk et al. 2007
38	*Chlorophanes spiza*	CytB	Smith et al. 2014
39	*Chlorospingus flavopectus* *Chlorospingus semifuscus* *Chlorospingus inornatus* *Chlorospingus tacarcunae*	ATP8-ATP6	Garcia-Moreno et al. 2004, Weir et al. 2008, Bonaccorso et al. 2008
40	*Chrysomus icterocephalus*	ND2	Cadena et al. 2011
41	*Cinclodes fuscus* *Cinclodes olrogi* *Cinclodes oustaleti* *Cinclodes comechingonus* *Cinclodes antarcticus* *Cindlodes albidiventris* *Cinclodes albiventris*	ND3	Chesser 2004, Sanin et al. 2009
42	*Colonia colonus*	CytB	Smith et al. 2014
43	*Corvus corax* *Corvus cryptoleucus*	CytB	Omland et al. 2000

APPENDIX D CONT.

44	*Cranioleuca antisiensis* *Cranioleuca baroni*	ND2	Seeholzer and Brumfield
45	*Cyanerpes caeruleus*	CytB	Smith et al. 2014
46	*Cyanocitta stelleri*	ND2	Klicka et al.
47	*Cyanocompsa cyanoides*	CytB	Bryson et al. 2014
48	*Cyanolyca viridicyanus* *Cyanolyca turcosa* *Cyanolyca armillata*	ND2	Cuervo et al. 2013
49	*Cyclarhis gujanensis* *Cyclarhis nigrirostris*	ND2	Smith et al. 2012; Klicka et al.
50	*Cymbilaimus lineatus*	CytB	Smith et al. 2014
51	*Dendragapus fuliginosus* *Dendragapus obscurus*	CR	Barrowclough et al. 2004
52	*Dendrocincla fuliginosa* *Dendrocincla anabatina*	CytB	Smith et al. 2014
53	*Dendrocolaptes platyrostris*	CytB	Cabanne et al. 2011
54	*Diglossa caerulescens*	ND2	Cuervo et al. 2013
55	*Diglossa cyanea*	ND2	Cuervo et al. 2013
56	*Drymophila caudata* *Drymophila klagesi* *Drymophila hellmayri* *Drymophila striaticeps*	ND2	Isler et al. 2012; Cuervo et al. 2013
57	*Drymophila devillei*	ND2	Bates et al. 1999
58	*Dubusia taeniata*	ND2	Cuervo et al. 2013
59	*Empidonax difficilis* *Empidonax occidentalis*	ND2	Klicka et al.
60	*Empidonax flavescens*	ND2	Klicka et al.
61	*Empidonax traillii*	CytB	Paxton 2000
62	*Forpus coelestis* *Forpus conspicillatus* *Forpus xanthops* *Forpus passerinus* *Forpus xanthopterygius*	ND2-CytB	Smith et al. 2013
63	*Geothlypis tolmiei*	ND2	Klicka et al.
64	*Geothlypis trichas* *Geothlypis beldingi* *Geothlypis nelsoni* *Geothlypis flavovelata*	ND2	Klicka et al.
65	*Glyphorynchus spirurus*	CytB	Smith et al. 2014
66	*Habia fuscicauda*	ND2	Klicka et al.
67	*Habia rubica*	ND2	Klicka et al.
68	*Hellmayrea gularis*	ND2	Cuervo et al. 2013

APPENDIX D CONT.

69	*Henicorhina leucophrys* *Henicorhina negreti*	ATP8-ATP6	Klicka et al.
70	*Henicorhina leucoptera*	ATP8-ATP6	Klicka et al.
71	*Henicorhina leucosticta*	CytB	Smith et al. 2014
72	*Hirundo rustica*	ND2	Dor et al. 2010
73	*Hylophilus ochraceiceps*	ND2Smith et al. 2012	
74	*Hylophylax naevioides* *Hylophylax naevius*	ND2	Fernandes et al. 2014
75	*Hypocnemis cantator* *Hypocnemis flavescens* *Hypocnemis peruviana* *Hypocnemis subflava* *Hypocnemis ochrogyna* *Hypocnemis striata* *Hypocnemis rondoni*	ND2	Bates et al. 2000, Tobias et al. 2008, Naka et al. 2012
76	*Icterus galbula* *Icterus abeillei*	CytB	Kondo et al. 2004
77	*Icterus pustulatus*	CytB	Cortes-Rodriguez et al. 2008a
78	*Lampornis amethystinus*	CR	Cortes-Rodriguez et al. 2008b
79	*Lepidocolaptes affinis*	ND2	Arbeláez-Cortes et al. 2010
80	*Lepidocolaptes lacrymiger*	ND2	Cuervo et al. 2013
81	*Lepidothrix coronata*	CytB	Smith et al. 2014
82	*Margarornis squamiger*	ND2	Cuervo et al. 2013
83	*Mecocerculus leucophrys*	ND2	Cuervo et al. 2013
84	*Melanerpes formicivorus*	CytB	Honey-Escandon et al. 2008
85	*Melozone fusca*	CR	Zink et al. 2001
86	*Microbates cinereiventris* *Microbates collaris*	ND2	Naka et al. 2012; Smith et al.; Smith et al. 2012
87	*Microcerculus marginatus*	CytB	Smith et al. 2014
88	*Mionectes oleagineus* *Mionectes rufiventris* *Mionectes macconnelli*	ND2	Miller at al. 2007
89	*Mionectes striaticollis*	ND2	Cuervo et al. 2013
90	*Momotus mexicanus*	ND2	Arbelaez-Cortes et al. 2013
91	*Myadestes occidentalis*	ND2	Barber and Klicka 2010
92	*Myioborus miniatus*	ND2	Pérez-Emán 2005, Perez-Eman et al. 2010, Klicka et al.
93	*Myiothlypis coronata*	ND2	Cuervo et al. 2013
94	*Myiothlypis fulvicauda* *Myiothlypis rivularis*	ATP8-ATP7	Lovette 2004
95	*Myiothlypis leucoblephara*	CytB	Batalha-Filho et al. 2012

APPENDIX D CONT.

96	*Myiothlypis luteoviridis*	ND2	Cuervo et al. 2013
97	*Myrmeciza exsul*	ND2	Miller et al. 2010
98	*Myrmeciza hemimelaena*	CytB	Fernandes et al. 2012
99	*Myrmeciza loricata* *Myrmeciza squamosa*	ND2	Amaral et al. 2013
100	*Myrmotherula axillaris*	CytB	Smith et al. 2014
101	*Myrmotherula schisticolor*	ND2	Cuervo et al. 2013
102	*Nucifraga columbiana*	ND2	Dohms and Burg 2013
103	*Ochthoeca cinnamomeiventris*	ND2	Cuervo et al. 2013
104	*Oreothlypis superciliosa*	ND2	Barber and Klicka 2010
105	*Passerculus sandwichensis*	ND2	Zink et al. 2005
106	*Passerina ciris*	ND2	Herr et al. 2011
107	*Passerina leclancherii*	ND2	Arbelaez-Cortes et al. 2013
108	*Perisoreus canadensis*	ND2	van Els et al. 2012
109	*Petrochelidon fulva*	CytB	Kirchman et al. 2000
110	*Phaethornis guy*	ND2	Cuervo et al. 2013
111	*Pheucticus melanocephalus*	ND2	van Els et al. 2014
112	*Phrygilus fruticeti*	COI	Campagna et al. 2011
113	*Piaya cayana*	CytB	Smith et al. 2014
114	*Picoides dorsalis*	CytB	Zink et al. 2002
115	*Picoides pubescens*	CR-ATP8-ATP6	Pulgarin and Burg 2012
116	*Picoides villosus*	ND2	Klicka et al. 2011
117	*Pinicola enucleator*	ND2	Drovetski et al. 2010
118	*Pipilo erythrophthalmus* *Pipilo maculatus*	ND2	Klicka et al.
119	*Pipreola riefferii* *Pipreola intermedia*	ND2	Cuervo et al. 2013
120	*Piranga flava*	ND2	Klicka et al.
121	*Piranga ludoviciana*	CytB	Klicka et al.
122	*Poecile atricapillus*	ND6	Pravosudov et al. 2012
123	*Poecile gambeli*	ND2	Spellman et al. 2007
124	*Polioptila albiloris* *Polioptila nigriceps* *Polioptila melanura* *Polioptila californica* *Polioptila plumbea* *Polioptila caerulea* *Polioptila dumicola* *Polioptila lactea*	ND2	Smith et al.; Smith et al. 2012; Zink et al. 2013

APPENDIX D CONT.

125	*Premnoplex tatei* *Premnoplex brunnescens*	ND2	Valderrama et al. 2014
126	*Psaltriparus minimus*	ND2	Klicka et al.
127	*Pyriglena leucoptera* *Pyriglena atra*	ND2	Maldonado-Coelho et al. 2013
128	*Pyrrhomyias cinnamomeus*	ND2	Cuervo et al. 2013
129	*Querula purpurata*	CytB	Smith et al. 2014
130	*Quiscalus mexicanus* *Quiscalus major*	ND2	DaCosta et al. 2008
131	*Rallus longirostris* *Rallus elegans*	ND2	Maley and Brumfield 2013
132	*Ramphocaenus melanurus*	ND2	Smith et al. 2012; Smith et al.
133	*Regulus calendula*	ATP8-ATP6	Klicka et al.
134	*Regulus satrapa*	ATP8-ATP6	Klicka et al.
135	*Saltator coerulescens* *Saltator similis* *Saltator striatipectus*	ND2	Chaves et al. 2013
136	*Saltator grossus* *Saltator cinctus* *Saltator aurantiirostris*	ND2	Chaves et al. 2013
137	*Saltator maximus*	ND2	Chaves et al. 2013
138	*Schiffornis turdina*	CytB	Smith et al. 2014
139	*Schiffornis virescens*	CR	Cabanne et al. 2012
140	*Sclerurus mexicanus* *Sclerurus rufigularis*	CytB	Smith et al. 2014
141	*Sclerurus scansor*	ND2	d'Horta et al. 2013
142	*Selasphorus platycercus*	CR	Malpica and Ornelas 2014
143	*Setophaga coronata*	ATP8-ATP6	Klicka et al.
144	*Setophaga dominica*	CR	McKay 2009
145	*Setophaga graciae*	ND2	Klicka et al.
146	*Setophaga petechia*	CR	Milot et al. 2000
147	*Setophaga ruticilla*	CR	Colbeck et al. 2008
148	*Setophaga striata*	CR	Ralston and Kirchman 2012
149	*Sialia mexicana* *Sialia currucoides* *Sialia sialis*	ATP8-ATP6	Klicka et al.
150	*Sitta carolinensis*	ND2	Spellman and Klicka 2007
151	*Spizella passerina*	CR	Mila et al. 2006
152	*Strix occidentalis*	CR	Barrowclough et al. 2011

APPENDIX D CONT.

153	*Strix varia*	CR	Barrowclough et al. 2011
154	*Sturnella magna* *Sturnella neglecta*	ND2	Barker et al. 2008
155	*Synallaxis azarae* *Synallaxis courseni*	ND2	Cuervo et al. 2013
156	*Tachycineta bicolor*	ND2	Stenzler et al. 2009
157	*Tangara cyanicollis*	CytB	Smith et al. 2014
158	*Tangara gyrola*	CytB	Smith et al. 2014
159	*Tangara vassorii*	ND2	Cuervo et al. 2013
160	*Tersina viridis*	CytB	Smith et al. 2014
161	*Tityra semifasciata*	CytB	Smith et al. 2014
162	*Toxostoma curvirostre*	ND2	Rojas-Soto et al. 2007
163	*Toxostoma redivivum*	CytB	Sgariglia and Burns 2003
164	*Troglodytes aedon*	ND2	Chaves et al.
165	*Troglodytes hiemalis* *Troglodytes pacificus*	ND2	Drovetski et al. 2004
166	*Trogon rufus*	CytB	Smith et al. 2014
167	*Vireo atricapilla*	ND2	Zink et al. 2010
168	*Vireo gilvus*	ND2	Klicka et al.
169	*Vireo huttoni*	ND2	Klicka et al.
170	*Vireo solitarius* *Vireo plumbeous* *Vireo cassinii*	ND2	Klicka et al.
171	*Xenopipo atronitens*	ND2	Capurucho et al. 2013
172	*Xenops minutus*	CytB	Smith et al. 2014
173	*Xiphorhynchus elegans* *Xiphorhynchus spixii*	CytB	Aleixo 2004
174	*Xiphorhynchus fuscus*	ND2	Cabanne et al. 2008
175	*Xiphorhynchus ocellatus* *Xiphorhynchus pardalotus*	ND2	Sousa-Neves et al. 2013
176	*Xiphorhynchus triangularis*	ND2	Cuervo et al. 2013
177	*Zonotrichia capensis*	CR	Lougheed et al. 2013

APPENDIX D CONT.

Supplemental References

S. Reddy, What's missing from avian global diversification analyses? *Mol. Phylo. Evol.* **77**, 159-165 (2014).

B. T. Smith, J. Klicka, The profound influence of the Late Pliocene Panamanian uplift on the exchange, diversification, and distribution of New World birds. *Ecography* **33**, 333-342 (2010).

J. T. Weir, D. Schluter, Calibrating the avian molecular clock. *Molecular Ecology* **17**, 2321-2328.

K. Tamura, G. Stecher, D. Peterson, A. Filipski, S. Kumar, MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol. Biol. Evol.* **30**, 2725-2729.

A. Rambaut, A. J. Drummond, Tracer v.1.5. Available at http://tree.bio.ed.ac.uk/software/tracer (2007).

D. Rabosky, K. Schliep. Laser: Likelihood analysis of speciation/extinction rates from phylogenies. R package version 2.4-1. Available at http://CRAN.R-project.org/package=laser (2013).

S. J. Hackett *et al.*, A phylogenomic study of birds reveals their evolutionary history. *Science* **320**, 1763-1768 (2008).

P. G. P. Ericson, Evolution of terrestrial birds in three continents: Biogeography and parallel radiations. *J. Biogeogr.* **39**, 813-824 (2012).

APPENDIX D CONT.

D. L. Rabosky *et al.* Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. *Nature Comm.* **4**, 1958 (2013).

P. Brito, S. V. Edwards, Multilocus phylogeography and phylogenetics using sequence-based markers. *Genetica* **135**, 439-455 (2009).

Harvey and Brumfield 2015

R. J. Elshire, J. C. Glaubitz, Q. Sun, J. A. Poland, K. Kawamoto, E. S. Buckler, S. E. Mitchell, A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* **6**, e19379 (2011).

P. J. Bradbury, Z. Zhang, D. E. Kroon, T. M. Casstevens, Y. Ramdoss, E. S. Buckler, TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633-2635 (2007).

M. Lynch, Estimation of allele frequencies from high-coverage genome-sequencing projects. *Genetics* **182**, 295-301 (2009).

T. A. White, S. E. Perkins, G. Heckel, J. B. Searle. Adaptive evolution during an ongoing range expansion: The invasive bank vole (*Myodes glareolus*) in Ireland. *Mol. Ecol.* **22**, 2971-2985.

J. K. Pritchard, M. Stephens, P. Donnelly, Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959 (2000).

S. Palumbi, R. F. Cipriano, M. P. Hare, Predicting nuclear gene coalescence from mitochondrial data: the three-times rule. *Evolution* **55**, 859-868 (2001).

American Ornithologists' Union (AOU), *Check-list of North American Birds, 7th Edition* (American Ornithologists' Union, Washington, D. C., 1998).

APPENDIX D CONT.

R. T. Chesser *et al*., Fifth-fourth supplement to the American Ornithologists' Union Check-List of North American Birds. *Auk* **130**, 558-571 (2013).

J. V. Remsen Jr. *et al.*, A classification of the bird species of South America. Available at http://www.museum.lsu.edu/~Remsen/SACCBaseline.html (2014).

T. Stadler, D. L. Rabosky, R. E. Ricklefs, F. Bokma, On age and species richness in higher taxa. *Am. Nat*. **184**, 447-455 (2014).

M. Dynesius, R. Jansson, Persistence of within-species lineages: a neglected control of speciation rates. *Evolution* **68**, 923-934 (2014).

J. A. Chavez, J. P. Pollinger, T. B. Smith, G. LeBuhn, The role of geography and ecology in shaping the phylogeography of the Speckled Hummingbird (*Adelomyia melanogenys*) in Ecuador. *Mol. Phylo. Evol.* **43**, 795-807 (2007).

J. A. Chavez, J. T. Weir, T. B. Smith, Diversification in *Adelomyia* hummingbirds follows Andean uplift. *Molecular Ecology* **20**, 4564-4576 (2011).

F. K. Barker, M. K. Benesh, A. J. Vandergon, S. M. Lanyon, Contrasting evolutionary dynamics and information content of the avian mitochondrial control region and ND2 gene. *PLoS One* **7**, e46403 (2012).

M. J. Lelevier, thesis, University of Alaska (2011).

T. J. Wenner, M. A. Russello, T. F. Wright, Cryptic species in a Neotropical parrot: Genetic variation within the *Amazona farinosa* complex and its conservation implications. *Cons. Genet.* **13**, 1427-1432 (2012).

J. R. Eberhard, E. Bermingham, Phylogeny and biogeography of the *Amazona ochrocephala* (Aves: Psittacidae) complex. *Auk* **121**, 318-332 (2004).

C. C. Ribas, E. S. Tavares, C. Yoshihara, C. Y. Miyaki, Phylogeny and biogeography of Yellow-headed and Blue-fronted parrots (*Amazona ochrocephala* and *Amazona aestiva*) with special reference to the South American taxa. *Ibis* **149**, 564-574 (2007).

A. M. Cuervo, Dissertation, Louisiana State University (2013).

K. G. McCracken, W. P. Johnson, F. H. Sheldon, Molecular population genetics, phylogeography, and conservation biology of the Mottled Duck (*Anas fulvigula*). *Cons. Genet.* **2**, 87-102 (2001).

J. L. Peters, K. E. Omland, Population structure and mitochondrial polyphyly in North American Gadwalls (*Anas strepera*). *Auk* **124**, 444-462 (2007).

J. E. McCormack, J. Heled, K. S. Delaney, A. T. Peterson, L. L. Knowles, Calibrating divergence times on species trees versus gene trees: Implications for speciation history of *Aphelocoma* jays. *Evolution* **65**, 184-202 (2011).

J. E. McCormack, A. T. Peterson, E. Bonaccorso, T. B. Smith, Speciation in the highlands of Mexico: Genetic and phenotypic divergence in the Mexican jay (*Aphelocoma ultramarina*). *Mol. Ecol.* **17**, 2505-2521 (2008).

C. C. Ribas, C. Y. Miyaki, Molecular systematics in *Aratinga* parakeets: Species limits and historical biogeography in the 'solstitialis' group and the systematic position of *Nandayus nenday*. *Mol. Phylo. Evol.* **30**, 663-675 (2004).

A. G. Navarro-Sigüenza, A. T. Peterson, A. Nyari, G. M. García-Deras, J. García-Moreno, Phylogeography of the *Buarremon* brush-finch complex (Aves, Emberizidae) in Mesoamerica. *Mol. Phylo. Evol*. **47**, 21-35 (2008).

APPENDIX D CONT.

C. D. Cadena, A. M. Cuervo, Molecules, ecology, morphology, and songs in concert: how many species is *Arremon torquatus* (Aves: Emberizidae)? *Biol. J. Linn. Soc.* **99**, 152-176 (2010).

C. Cicero, M. S. Koo, The role of niche divergence and phenotypic adaptation in promoting lineage diversification in the Sage Sparrow (*Artemisiospiza belli*, Aves: Emberizidae). *Biol. J. Linn. Soc.* **107**, 332-354 (2012).

F. Puebla-Olivares *et al.*, Speciation in the Emerald Toucanet (*Aulacorhynchus prasinus*) complex. *Auk* **125**, 39-50 (2008).

C. Cicero, Barriers to sympatry between avian sibling species (Paridae: *Baeolophus*) in local secondary contact. *Evolution* **58**, 1573-1587 (2004).

B. R. Barber, J. Klicka, Two pulses of diversification across the Isthmus of Tehuantepec in a montane Mexican bird fauna. *Proc. R. Soc. B* **282**, 2675-2681 (2010).

S. T. Vilaça, F. R. Santos, Biogeographic history of the species complex *Basileuterus culicivorus* (Aves, Parulidae) in the Neotropics. *Mol. Phylo. Evol.* **57**, 585-597 (2010).

N. Gutiérrez-Pinto *et al.*, Non-monophyly and deep genetic differentiation across low-elevation barriers in a Neotropical montane bird (*Basileuterus tristriatus*; Aves: Parulidae). *Mol. Phylo. Evol.* **64**, 156-165 (2012).

C. L. Pruett, K. Winker, Biological impacts of climatic change on a Beringian endemic: Cryptic refugia in the establishment and differentiation of the Rock Sandpiper (*Calidris ptilocnemis*). *Climatic Change* **68**, 219-240 (2005).

APPENDIX D CONT.

R. M. Zink, A. E. Kessen, T. V. Line, R. C. Blackwell-Rago, Comparative phylogeography of some aridland bird species. *Condor* **103**, 1-10 (2001).

H. Vásquez-Miranda, A. G. Navarro-Sigüenza, K. E. Omland, Phylogeography of the Rufous-naped Wren (Campylorhynchus rufinucha): Speciation and hybridization in Mesoamerica. *Auk* **126**, 765-778 (2009).

M. A. González, J. R. Eberhard, I. J. Lovette, S. L. Olson, E. Bermingham, Mitochondrial DNA phylogeography of the Bay Wren (Troglodytidae: *Thryothorus nigricapillus*) complex. *Condor* **105**, 228-238 (2003).

G. F. Seeholzer, B. M. Winger, M. G. Harvey, J. D. Weckstein, A new species of barbet (Capitonidae: *Capito*) from the Cerros del Sira, Ucayali, Peru. *Auk* **129**, 551-559 (2011).

M. Kimura *et al.*, Phylogeographical approaches to assessing demographic connectivity between breeding and overwintering regions in a Nearctic-Neotropical warbler (*Wilsonia pusilla*). *Mol. Ecol.* **11**, 1605-1616 (2002).

A. O. Barrera-Guzmán, B. Milá, L. A. Sánchez-González, A. G. Navarro-Sigüenza, Speciation in an avian complex endemic to the mountains of Middle America (*Ergaticus*, Aves: Parulidae). *Mol. Phylo. Evol.* **62**, 907-920 (2012).

B. T. Smith *et al.* The role of historical and contemporary processes on phylogeographic structure and genetic diversity in the Northern Cardinal, *Cardinalis cardinalis*. *BMC Evol. Biol.* **11**, 136 (2011).

APPENDIX D CONT.

J. D. Manthey, J. Klicka, G. M. Spellman, Cryptic diversity in a widespread North American songbird: Phylogeography of the Brown Creeper (*Certhia americana*). *Mol. Phylo. Evol.* **58**, 502-512 (2011).

K. J. Burns, D. N. Barhoum, Population-level history of the Wrentit (*Chamaea fasciata*): Implications for comparative phylogeography in the California Floristic Province. *Mol. Phylo. Evol.* **38**, 117-129 (2006).

S. J. Oyler-McCance, J. St. John, F. L. Knopf, T. W. Quinn, Population genetic analysis of Mountain Plover using mitochondrial DNA sequence data. *Condor* **107**, 353-362 (2005).

W. C. Funk, T. D. Mullins, S. M. Haig, Conservation genetics of Snowy Plovers (*Charadrius alexandrinus*) in the Western Hemisphere: Population genetic structure and delineation of a new subspecies. *Cons. Genet.* **8**, 1287-1309 (2007).

J. García-Moreno, A. G. Navarro-Sigüenza, A. T. Peterson, L. A. Sánchez-González, Genetic variation coincides with geographic structure in the Common Bush-Tanager (*Chlorospingus ophthalmicus*) complex from Mexico. *Mol. Phylo. Evol.* **33**, 186-196 (2004).

J. T. Weir, E. Bermingham, M. J. Miller, J. Klicka, M. A. González, Phylogeography of a morphologically diverse Neotropical montane species, the Common Bush-Tanager (*Chlorospingus ophthalmicus*). *Mol. Phylo. Evol.* **47**, 650-664 (2008).

E. Bonaccorso, A. G. Navarro-Sigüenza, L. A. Sánchez-González, A. T. Peterson, J. García-Moreno, Genetic differentiation of the *Chlorospingus ophthalmicus* complex in Mexico and Central America. *J. Avian. Biol.* **39**, 311-321 (2008).

C. D. Cadena, N. Gutiérrez-Pinto, N. Dávila, R. T. Chesser, No population genetic structure in a widespread aquatic songbird from the Neotropics. *Mol. Phylo. Evol.* **58**, 540-545 (2011).

R. T. Chesser, Systematics, evolution, and biogeography of the South American ovenbird genus *Cinclodes*. *Auk* **121**, 752-766 (2004).

C. Sanín *et al*., Paraphyly of *Cinclodes fuscus* (Aves: Passeriformes: Furnariidae): Implications for taxonomy and biogeography. *Mol. Phylo. Evol.* **53**, 547-555 (2009).

K. E. Omland, C. L. Tarr, W. I. Boarman, J. M. Marzluff, R. C. Fleischer, Cryptic genetic variation and paraphyly in ravens. *Proc. R. Soc. B* **1461**, 2475-2482 (2000).

R. W. Bryson *et al.*, Diversification across the New World within the "blue" Cardinalids (Aves: Cardinalidae). *J. Biogeogr.* **41**, 587-599 (2014).

B. T. Smith, A. Amei, J. Klicka, Evaluating the role of contracting and expanding rainforest in initiating cycles of speciation across the Isthmus of Panama, *Proc. R. Soc. B* **279**, 3520-3526 (2012).

G. F. Barrowclough, J. G. Groth, L. A. Mertz, R. J. Gutiérrez, Phylogeographic structure, gene flow, and species status in Blue Grouse (*Dendragapus obscurus*). *Mol. Ecol.* **7**, 1911-1922 (2004).

G. S. Cabanne, F. M. D'Horta, D. Meyer, J. Silva, C. Y. Miyaki , Evolution of *Dendrocolaptes platyrostris* (Aves: Furnariidae) between the South American open vegetation corridor and the Atlantic Forest. *Biol. J. Linn. Soc*. **103**, 801-820 (2011).

APPENDIX D CONT.

M. L. Isler, A. M. Cuervo, G. A. Bravo, R. T. Brumfield, An integrative approach to species-level systematics reveals the depth of diversification in an Andean Thamnophilid, the Long-tailed Antbird. *Auk* **114**, 571-583 (2012).

J. M. Bates, S. J. Hackett, J. M. Goerck, High levels of mitochondrial DNA differentiation in two lineages of antbirds (*Drymophila* and *Hypocnemis*). *Auk* **116**, 1093-1106 (1999).

E. H. Paxton, Dissertation, Northern Arizona University (2000).

B. T. Smith, C. C. Ribas, B. M. Whitney, B. E. Hernández-Baños, J. Klicka, Identifying biases at different spatial and temporal scales of diversification: A case study in the Neotropical parrotlet genus *Forpus*. *Mol. Ecol*. **22**, 483-494 (2013).

R. Dor, R. J. Safran, F. H. Sheldon, D. W. Winkler, I. J. Lovette, Phylogeny of the genus *Hirundo* and the Barn Swallow subspecies complex. *Mol. Phylo. Evol.* **56**, 409-418 (2010).

A. M. Fernandes, M. Wink, C. H. Sardelli, A. Aleixo, Multiple speciation across the Andes and throughout Amazonia: The case of the Spot-backed Antbird species complex (*Hylophylas naevius/Hylophylax naevioides*). *J. Biogeogr.* **41**, 1094-1104 (2014).

J. A. Tobias, J. M. Bates, S. J. Hackett, N. Seddon, Comment on "The latitudinal gradient in recent speciation and extinction rates of birds and mammals". *Science* **319**, 901c (2008).

L. N. Naka, C. L. Bechtoldt, L. M. P. Henriques, R. T. Brumfield, The role of physical barriers in the location of avian suture zones in the Guiana Shield, northern Amazonia. *Am. Nat.* **179**, E115-E132 (2012).

B. Kondo, J. M. Baker, K. E. Omland, Recent speciation between the Baltimore Oriole and the Black-backed Oriole. *Condor* **106**, 674-680 (2004).

N. Cortes-Rodríguez, B. E. Hernández-Baños, A. G. Navarro-Sigüenza, K. E. Omland, Geographic variation and genetic structure in the Streak-backed Oriole: Low mitochondrial DNA differentiation reveals recent divergence. *Condor* **110**, 729-739 (2008).

N. Cortes-Rodríguez, B. E. Hernández-Baños, A. G. Navarro-Sigüenza, A. T. Peterson, J. García-Moreno, Phylogeography and population genetics of the Amethyst-throated Hummingbird (*Lampornis amethystinus*). *Mol. Phylo. Evol.* **48**, 1-11 (2008).

E. Arbeláez-Cortés, Á. S. Nyári, A. G. Navarro-Sigüenza, The differential effect of lowlands on the phylogeographic pattern of a Mesoamerican montane species (*Lepidocolaptes affinis*, Aves: Furnariidae). *Mol. Phylo. Evol.* **57**, 658-668 (2010).

M. Honey-Escandón, B. E. Hernández-Baños, A. G. Navarro-Sigüenza, H. Benítez-Díaz, A. T. Peterson, Phylogeographic patterns of differentiation in the Acorn Woodpecker. *Wilson J. Ornith.* **120**, 478-493 (2008).

M. J. Miller *et al*., Out of Amazonia again and again: Episodic crossing of the Andes promotes diversification in a lowland forest flycatcher. *Proc. R. Soc. B* **275**, 1133-1142 (2008).

APPENDIX D CONT.

E. Arbeláez-Cortés, B. Milá, A. G. Navarro-Sigüenza, Multilocus analysis of intraspecific differentiation in three endemic bird species from northern Neotropical dry forest. *Mol. Phylo. Evol.* **70**, 362-377 (2014).

J. L. Pérez-Emán, Molecular phylogenetics and biogeography of the Neotropical redstarts (*Myioborus*; Aves, Parulinae). Mol. Phylo. Evol. 37, 511-528 (2005).

J. L. Pérez-Emán, R. L. Mumme, P. G. Jablonski, Phylogeography and adaptive plumage evolution in Central American subspecies of the Slate-throated Redstart (*Myioborus miniatus*). *Ornith. Monogr.* **1**, 90-102 (2010).

I. J. Lovette, Molecular phylogeny and plumage signal evolution in a trans Andean and circum Amazonian avian species complex. *Mol. Phylo. Evol.* **32**, 512-523 (2004).

H. Batalha-Filho, G. S. Cabanne, C. Y. Miyaki, Phylogeography of an Atlantic forest passerine reveals demographic stability through the last glacial maximum. *Mol. Phylo. Evol*. **65**, 892-902 (2012).

M. J. Miller, E. Bermingham, J. Klicka, P. Escalante, K. Winker, Neotropical birds show a humped distribution of within-population genetic diversity along a latitudinal transect. *Ecology Letters* **13**, 576-586 (2010).

A. M. Fernandes, M. Wink, A. Aleixo, Phylogeography of the Chestnut-tailed Antbird (Myrmeciza hemimelaena) clarifies the role of rivers in Amazonian biogeography. *J. Biogeogr.* **39**, 1524-1535 (2012).

F. Raposo do Amaral, P. K. Albers, S. V. Edwards, C. Y. Miyaki, Multilocus tests of Pleistocene refugia and ancient divergence in a pair of Atlantic Forest antbirds (*Myrmeciza*). *Mol. Ecol.* **22**, 3996-4013 (2013).

APPENDIX D CONT.

K. M. Dohms, T. M. Burg, Molecular markers reveal limited population genetic structure in a North American Corvid, Clark's Nutcracker (*Nucifraga columbiana*). *PLoS One* **8**, e79621 (2013).

R. M. Zink *et al.*, Mitochondrial DNA variation, species limits, and rapid evolution of plumage coloration and size in the Savannah Sparrow. *Condor* **107**, 21-28 (2005).

C. A. Herr, P. W. Sykes Jr., J. Klicka, Phylogeography of a vanishing North American songbird: the Painted Bunting (*Passerina ciris*). *Conserv. Genet*. **12**, 1395-1410 (2011).

P. van Els, C. Cicero, J. Klicka, High latitudes and high genetic diversity: Phylogeography of a widespread boreal bird, the Gray Jay (*Perisoreus canadensis*). *Mol. Phylo. Evol.* **63**, 456-465 (2012).

J. J. Kirchman, L. A. Whittingham, F. H. Sheldon, Relationships among Cave Swallow populations (*Petrochelidon fulva*) determined by comparisons of microsatellite and cytochrome b data. *Mol. Phylo. Evol.* **14**, 107-121 (2000).

L. Campagna *et al.*, A molecular phylogeny of the Sierra-Finches (*Phrygilus*, Passeriformes): Extreme polyphyly in a group of Andean specialists. *Mol. Phylo. Evol.* **61**, 521-533 (2011).

R. M. Zink, S. Rohwer, S. Drovetski, R. C. Blackwell-Rago, S. L. Farrell, Holarctic phylogeography and species limits of Three-toed Woodpeckers. *Condor* **104**, 167-170 (2002).

APPENDIX D CONT.

P. C. Pulgarín-R., T. M. Burg, Genetic signals of demographic expansion in Downy Woodpecker (*Picoides pubescens*) after the last North American glacial maximum. *PLoS One* **7**, e40412 (2012).

J. Klicka, G. M. Spellman, K. Winker, V. Chua, B. T. Smith, A phylogeographic and population genetic analysis of a widespread, sedentary North American bird: The Hairy Woodpecker (*Picoides villosus*). *Auk* **128**, 346-362 (2011).

S. V. Drovetski, R. M. Zink, P. G. P. Ericson, I. V. Fadeev, A multilocus study of Pine Grosbeak phylogeography supports the pattern of greater intercontinental divergence in Holarctic boreal forest birds than in birds inhabiting other high-latitude habitats. *J. Biogeogr*. **37**, 696-706 (2010).

V. V. Pravosudov *et al.*, Population genetic structure and its implications for adaptive variation in memory and the hippocampus on a continental scale in food-caching Black-capped Chickadees. *Mol. Ecol.* **21**, 4486-4497 (2012).

G. M. Spellman, B. Riddle, J. Klicka, Phylogeography of the Mountain Chickadee (*Poecile gambeli*): Diversification, introgression, and expansion in response to Quaternary climate change. *Mol. Ecol*. **16**, 1055-1068 (2007).

R. M. Zink, J. G. Groth, H. Vásquez-Miranda, G. F. Barrowclough, Phylogeography of the California Gnatcatcher (Polioptila californica) using multilocus DNA sequences and ecological niche modeling: Implications for conservation. *Auk* **130**, 449-458 (2013).

E. Valderrama, J. L. Pérez-Emán, R T. Brumfield, A. M. Cuervo, C. D. Cadena, The influence of the complex topography and dynamic history of the montane

Neotropics on the evolutionary differentiation of a cloud forest bird (*Premnoplex brunnescens,* Furnariidae). *J. Biogeogr*. **41**, 1533-1546 (2014).

M. Maldonado-Coelho, J. G. Blake, L. F. Silveira, H. Batalha-Filho, R. E. Ricklefs, Rivers, refuges, and population divergence of fire-eye antbirds (*Pyriglena*) in the Amazon Basin. J. Evol. Biol. 26, 1090-1107 (2013).

J. M. DaCosta, W. Wehtje, J. Klicka, Historic genetic structuring and paraphyly within the Great-tailed Grackle. *Condor* **110**, 170-177 (2008).

J. M. Maley, R. T. Brumfield, Mitochondrial and next-generation sequence data used to infer phylogenetic relationships and species limits in the Clapper/King rail complex. *Condor* **115**, 316-329 (2013).

J. A. Chaves, J. R. Hidalgo, J. Klicka, Biogeography and evolutionary history of the Neotropical genus *Saltator* (Aves: Thraupini). *J. Biogeogr.* **40**, 2180-2190 (2013).

G. S. Cabanne, E. H. R. Sari, D. Meyer, F. R. Santos, C. Y. Miyaki, Matrilineal evidence for demographic range expansion, low diversity, and lack of phylogeographic structure in the Atlantic forest endemic Greenish Schiffornis *Schiffornis virescens* (Aves: Tityridaae). *J. Ornith.* **154**, 371-384 (2013).

F. M. d'Horta, A. M. Cuervo, C. C. Ribas, R. T. Brumfield, C. Y. Miyaki, Phylogeny and comparative phylogeography of *Sclerurus* (Aves: Furnariidae) reveal constant and cryptic diversification in an old radiation of rain forest understory specialists. *J. Biogeogr*. **40**, 37-49 (2013).

APPENDIX D CONT.

A. Malpica, J. F. Ornelas, Postglacial northward expansion and genetic differentiation between migratory and sedentary populations of the Broad-tailed Hummingbird (*Selasphorus platycercus*). *Mol. Ecol.* **23**, 435-452 (2014).

B. D. McKay, Evolutionary history suggests rapid differentiation in the Yellow-throated Warbler (*Dendroica dominica*). *J. Avian Biol.* **40**, 181-190 (2009).

E. H. Milot, H. L. Gibbs, K. A. Hobson, Phylogeography and genetic structure of northern populations of the Yellow Warbler (*Dendroica petechia*). *Mol. Ecol.* **9**, 667-681 (2000).

G. J. Colbeck, H. L. Gibbs, P. P. Marra, K. A. Hobson, M. S. Webster, Phylogeography of a widespread North American migratory songbird (*Setophaga ruticilla*). *J. Hered.* **99**, 453-463 (2008).

J. Ralston, J. J. Kirchman, Continent-scale genetic structure in a boreal forest migrant, the Blackpoll Warbler (*Setophaga striata*). *Auk* **129**, 467-478 (2012).

G. M. Spellman, J. Klicka, Phylogeography of the White-breasted Nuthatch (*Sitta carolinensis*): Diversification in North American pine and oak woodlands. *Mol. Ecol.* **16**, 1729-1740 (2007).

B. Milá, T. B. Smith, R. K. Wayne, Postglacial population expansion drives the evolution of long-distance migration in a songbird. *Evolution* **60**, 2403-2409 (2006).

G. F. Barrowclough, J. G. Groth, K. J. Odom, J. E. Lai, Phylogeography of Spotted Owl (*Strix occidentalis*): Species limits, multiple refugia, and range expansion. *Auk* **28**, 696-706 (2011).

APPENDIX D CONT.

F. K. Barker, A. J. Vandergon, S. M. Lanyon, Assessment of species limits among yellow-breasted meadowlarks (*Sturnella* spp.) using mitochondrial and sex-linked markers. *Auk* **4**, 869-879 (2008).

L. M. Stenzler *et al.*, Subtle edge-of-range genetic structuring in transcontinentally distributed North American Tree Swallows. *Condor* **111**, 470-478 (2009).

O. R. Rojas-Soto, A. Espinosa de los Monteros, R. M. Zink, Phylogeography and patterns of differentiation in the Curve-billed Thrasher. *Condor* **109**, 456-463 (2007).

E. A. Sgariglia, K. J. Burns, Phylogeography of the California Thrasher (*Toxostoma redivivum*) based on nested-clade analysis of mitochondrial-DNA variation. *Auk* **120**, 346-361 (2003).

S. V. Drovetski *et al*., Complex biogeographic history of a Holarctic passerine. *Proc. R. Soc. B* **271**, 545-551 (2004).

R. M. Zink, A. W. Jones, C. C. Farquhar, M. C. Westberg, J. I. Gonzalez Rojas, Comparison of molecular markers in the endangered Black-capped Vireo (*Vireo atricapilla*) and their interpretation in conservation. *Auk* **127**, 797-806 (2010).

VITA

Michael G. Harvey was born in 1985 in Nashua, New Hampshire to Janice and Robert (Neil) Harvey. He grew up exploring the woods, streams, swamps, and hills of southern New Hampshire alone or with his brother Dennis. After an intimate run-in with a Great Blue Heron intent on the same frogs he had been pursuing, Michael's passion for birds took off. The range of his forays increased as he grew older and began exploring the White Mountains, coastal New England, and eventually Florida, the Rockies, southeastern Arizona, and California. After briefly considering more practical pursuits, Michael enrolled at Cornell University to study biology and to learn from the many excellent researchers at the Cornell Lab of Ornithology. After course work on evolutionary biology and gaining experience in the Lab's molecular facilities under Irby J. Lovette, Michael became enthralled with the explanatory power of evolutionary theory. Also, Michael began finding ways to explore the fantastic avain diversity in Latin America, in part inspired by fellow Cornell students Pete Hosner, Mike Andersen, and Dan Lebbin. After years of planning, Michael, Ben Winger, and Glenn Seeholzer conducted their trip of a lifetime, the first ornithological expedition to the southern Cerros del Sira of Peru. Michael earned his Bachelor of Arts in Biology from Cornell in 2008, and decided that the best way to pursue his various passions was to apply to the doctoral program at Louisiana State University to study the evolutionary biology of Neotropical birds. He entered the program under the tutelage of Dr. Robb T. Brumfield in January of 2010.